U. S. DEPARTMENT OF COMMERCE
W. Averell Harriman, Secretary
NATIONAL BUREAU OF STANDARDS
E. U. Condon, Director

SEPTEMBER 1947
VOLUME • 39
NUMBER 3

JOURNAL OF RESEARCH

of the

NATIONAL BUREAU OF STANDARDS

Contents

RP		Page
1823.	Shunts and inductors for surge current measurements. John H. Park. (Price 25 cents)	191
1824.	Validity of the cosine-fourth-power law of illumination. Irvine C. Gardner. (Price 10 cents)	213
1825.	Acid-base reactions in benzene and other organic solvents: Behavior of bromphthalein magneta with different classes of organic basis. Marion Maclean Davis and Priscilla J. Schuhmann. (Price 20 cents)	221
1826.	Observations on the control of grain size in magnesium casting alloys. Vernon C. F. Holm and Alexander I. Krynitsky. (Price 10 cents)	265
1827.	Acceptance sampling by variables, with special reference to the case in which quality is measured by average or dispersion. John H. Curtiss. (Price 15 cents)	271
1828.	Freezing points of cobalt and nickel. Milton S. VanDusen and Andrew I. Dahl. (Price 10 cents)	291

For sale by the Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C. - Price 60 cents, $4.50 per year on subscription

UNITED STATES GOVERNMENT PRINTING OFFICE, WASHINGTON : 1947

U. S. Department of Commerce
National Bureau of Standards

Research Paper RP1823
Volume 39, September 1947

Part of the Journal of Research of the National Bureau of Standards

Shunts and Inductors for Surge-Current Measurements

By John H. Park

The special requirements that must be fulfilled by a shunt intended to be used in surge-current measurements are explained. A tubular shunt with coaxial potential leads that meets these requirements is described, and factors affecting its design are discussed. A theoretical derivation of the "skin effect" in this type of shunt at high frequencies is given in one of the appendices.

The advantages of using a mutual inductor for obtaining oscillograms of the rate of change of current during a surge are outlined, and several types of mutual inductors developed especially for this purpose are described. Theoretical derivations, given in the appendices, indicate that the concentric-tube mutual inductors described in this paper can be used to measure the high-frequency components of a current surge up to 70 megacycles with less than 10 percent error.

Several shunts and mutual inductors of the designs described in this paper were constructed for use in the high voltage laboratory at the National Bureau of Standards. Their complete description and oscillograms showing results obtained with them are included.

I. Introduction

Surge-current generators are widely used in high-voltage laboratories to simulate the heavy current component of a lightning discharge. They are regularly employed in testing protective devices that are intended to withstand lightning discharges and in determining the effects of heavy-current discharges on various materials and apparatus. In order to correlate the test results obtained in various laboratories, it is necessary to know the magnitude and wave-form of the surge current used in each test.

A generally accepted method of measuring heavy-current surges is to insert a shunt in the discharge circuit and to apply the voltage drop across this shunt, through a suitable cable, to the deflecting plates of a cathode-ray-oscillograph capable of recording the variation of this voltage with time. The over-all accuracy of this method is dependent upon the three components used in the measurement: the cathode-ray-oscillograph, the cable, and the shunt.

The development of the cathode-ray-oscillograph to its present state of efficacy has involved years of work of a large number of experimenters and theorists. At the present time there are several types of oscillographs having recording speeds and response accuracies suitable for surge measurements. In general they may be classified as (1) the high-voltage cold-cathode type with the recording film inside the vacuum chamber, (2) the hot-cathode type with the film inside the vacuum chamber, and (3) the hot-cathode sealed-tube type using a fluorescent screen photographed by means of an external camera. Results reported in the present paper were obtained using an oscillograph of the first type; however, with slight modifications in design constants the apparatus to be described would be equally suitable for use with other types of oscillographs.

In order to facilitate safety precautions and to allow one cathode-ray-oscillograph (CRO) to be used with any one of several surge generators, the oscillograph is usually located at some distance from the surge generator and its potential divider and/or shunt and a cable (40 to 100 ft. long) is used to connect them to the deflecting plates of the oscillograph. The purpose of this cable is to reproduce the voltage drop across the shunt, at the deflecting plates of the oscillograph without distortion, and it must not introduce other voltages by induction from nearby currents. The use of coaxial-type cables reduces inductive effects to

a minimum, and when constructed with insulating material of low dielectric loss, such as polyethylene, coaxial cables are capable of transmitting all components of a surge up to at least 100 megacycles/sec over the distances required without significant attenuation. The terminations of the cable, at the shunt and at the oscillograph, are both important.

Various types of shunts have been used for measuring heavy-current surges [1,2,3],[1] and long before surge generators were in general use, the design of alternating current shunts to insure low residual inductance was given considerable attention [4,5]. Thus, the design of the shunts now in use at this laboratory for surge-current measurements is based on experience with shunts used for measuring large alternating currents as well as on subsequent experience in the field of surge-current measurement. The present paper will discuss the design of surge-current shunts, describe those constructed, and indicate the results obtained with them.

This paper will also describe a new design of mutual inductor for measuring rate of change of current. Because it tends to magnify the high frequency components in current waves it has been useful in their measurement.

II. Factors To Be Considered in the Design of a Shunt

1. Choice of the Best Type

In choosing the best type of design for a shunt to be used in measuring surge currents the three most important requirements are (1) the effective impedance considered as a 4-terminal network must be constant over as great a range in frequency as possible, (2) inductive effects of parts of the current circuit, other than the shunt, upon the potential-lead circuit of the shunt should be a minimum, and (3) it must be possible to connect the sheath of the cable from the shunt to the CRO, to ground at or near the shunt without introducing induced voltages in the shunt potential circuit (this is desirable to minimize the flow of ground currents in the sheath of the cable which might cause induced voltage at the CRO plates).

In order to determine compliance with requirement number one, the configuration of the shunt should preferably be such that its inductance and skin effect can be computed for as great a range in frequency as possible; otherwise the comparison of various types of shunts would require the actual construction of a large number with subsequent experimental comparisons. Two general types lend themselves to computations: (a) the concentric tubular and (b) the flat strip. Illustrations of these, giving the methods of attaching potential leads, are shown in figure 1. Another type of shunt used by Bellaschi [1] in surge measurements consists of several twisted "hairpin" loops of resistance wire connected in parallel at the current terminals. This type of shunt has been found suitable for most surge-current measurements, but its inductance and skin effect cannot be accurately computed from theoretical formulas. An approximate theoretical computation by Brownlee [2] indicates that the twisted loop shunt would probably have a somewhat larger time constant than a flat strip shunt. Theoretical formulas for computing the inductance (assuming uniform distribution of current) and skin effect (for a limited frequency range) of the concentric tubular and flat strip shunts are

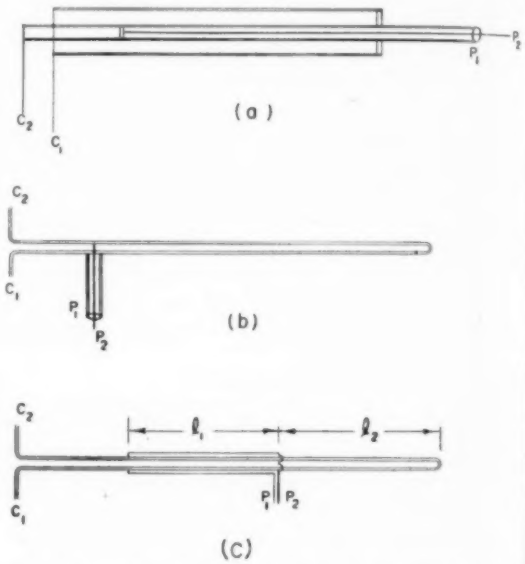


FIGURE 1.—*Three types of shunts whose inductance and skin effect can be computed.*

a, Concentric-tube shunt; *b*, flat-strip shunt; *c*, flat strip shunt with potential leads arranged to give minimum inductance.

[1] Figures in brackets indicate the literature references at the end of this paper.

readily available.[2 3] As suggested by Silsbee [4] a strip shunt with potential leads attached as shown in figure 1, c, could be designed so as to have zero effective inductance by a suitable choice of values of l_1 and l_2. However, this arrangement of potential leads prevents the shunt from complying with requirement (3) above, as will be explained later, so it is not considered in the following comparison of coaxial tubular vs. flat-strip shunts.

To compare the constancy of effective impedance as frequency changes, for the tubular and flat-strip shunts, two typical designs of shunt were assumed and their inductance and skin effect computed from the formulas referred to above. A concentric tubular shunt (fig. 1, a) in which the resistance material was a CuNi alloy ($\rho=22.9\text{x}10^{-6}$ ohm cm) tube whose outside diameter was 0.25 in. and wall thickness was 0.008 in. was chosen as one example because such a tube was readily available for use in constructing a shunt. For comparison purposes a strip shunt was chosen whose thickness was the same as the wall thickness of the tube, whose width was the same as the circumference of the tube, and whose separation between strips was taken to be 0.01 in. (The shunt length does not affect these computations provided it is several times the largest transverse dimension.) The time constants for the shunts as obtained using the theoretical formulas assuming uniform distribution of current were (1) for tubular shunt $L/R=-0.0365\text{x}10^{-6}$; and (2) for strip shunt $L/R=0.218\text{x}10^{-6}$. The skin effects at $f=1$ megacycle as obtained from formulas given by Silsbee were (1) for tubular shunt $R/R_{dc}=0.959$; and (2) for the strip shunt $R/R_{dc}=0.955$. It must be remembered that these results do not give an accurate evaluation of the constancy of effective impedance at frequencies over 1 megacycle, because at such frequencies the skin effect causes a nonuniform distribution of current density over the thickness of the tube or strip, for which the theoretical formulas used in the above computation do not hold—no published theoretical formulas good at these high frequencies were known to the author. However, the above results should serve their purpose as a comparison between the two types of shunts. The tubular shunt has a much lower time constant than the strip shunt, which means that its reactance will be negligible over a higher frequency range. The skin effect at 1 megacycle appears to be the same for both shunts, but the formula used in obtaining the skin effect for the strip shunt neglects edge effects that are known to considerably increase [5] the skin effect at higher frequencies. Thus it would appear that the tubular shunt, as it has no edge effects, would have less change in effective resistance at the higher frequencies than the strip shunt. A decrease in both inductance and skin effect could be obtained by using a thinner strip or a thinner walled tube, but this cannot be carried to extremes because of the heat capacity requirements of the shunt, which will be discussed later; any improvement made by this means in the strip shunt could probably also be made in the tubular shunt, though possibly not to the same extent. Thus, it would appear that either a coaxial tubular shunt or a flat strip shunt would satisfy requirement (1) (stated above), but that the tubular design would be preferable.

As indicated above, the readily available theoretical formulas are not capable of yielding the effective 4-terminal impedance of a strip or tubular shunt at frequencies of 1 megacycle or higher. Due to the axial symmetry of the tubular shunt, it is possible to develop a theoretical formula for calculating the impedance of such a shunt, taking into account the nonuniform distribution of current over the thickness of the tube (skin effect), which will hold for very high frequencies. The development of this theoretical formula (taken from hitherto unpublished notes of F. B. Silsbee) is given in appendix 1, and it permits the impedance of any tubular shunt to be computed for frequencies up to 10^9 or higher. The development of a similar formula for strip shunts would be much more difficult because of the lack of axial symmetry.

The inductive effects of current-carrying parts other than the shunt upon the potential lead circuit of the shunt (requirement 2), are dependent upon two factors: (a) the mutual inductance of the current carrying parts upon any loop in the potential circuit of the shunt, and (b) the maximum rate of change of current with time in the current-carrying part.

The maximum rate of change of current occurs on the front of the current wave, and it is some-

[2] See p. 400 of reference [4] for inductance of tubular shunt.

[3] See p. 81 of reference [5] for inductance of a flat-strip shunt; p. 86 for skin effect of flat strip, and p. 91 for skin effect of a tubular shunt.

[4] See p. 81 of reference [5].

[5] See reference [5], p. 86 and 87.

times taken to be the peak value of current divided by the time to peak. This is actually the average rate of change for the front of the wave, and any detailed study of the front of a current surge requires a more accurate basis for obtaining the maximum rate of change of current. Theoretically, the maximum rate of change of current for a capacitance discharging through an inductance and resistance occurs at the instant the circuit is closed and is equal to the voltage, E, to which the condenser was charged divided by the inductance, L, in the discharge circuit. The total inductance of the discharge circuit of a surge-current generator can be as low as 2 μh, and with a maximum voltage of 100 kv the maximum rate of change of current would be 5×10^{10} amp per second. Such a rate of change of current in the main current circuit would induce 50 v in the potential circuit of the shunt if the mutual inductance, M, between the two circuits were 0.001 μh. As a "pick-up loop" in the potential lead circuit of only 1 sq cm area, placed very close to a current-carrying part, may have a mutual inductance of 0.002 μh (giving a peak induced voltage of 100 v), the extreme importance of minimizing mutual inductance between the potential circuit of a shunt and all current-carrying parts becomes apparent. If the surge generator discharge circuit is fairly long so that its total inductance is more than 2 μh, the maximum rate of change of current will not be correspondingly reduced, because the stray capacitance across the discharge circuit will in effect short out the added inductance at the instant the surge is being initiated.

Referring to figure 1, a, the potential circuit of a tubular shunt may be considered as an extension of the coaxial cable from the CRO, ending in a direct short circuit from the central conductor to the sheath. Due to the axial symmetry of this arrangement the mutual inductance between the potential circuit and any current-carrying parts is the lowest obtainable. Even when the sheath of the cable is connected to ground and ground currents flow in the sheath of the cable and in the outer current return tube of the shunt itself, these ground currents will not induce voltage in the potential circuit of the shunt provided the density of such currents is symmetrical around the axis of the shunt. End effects or concentration of current along one side of the shunt axis can be largely eliminated by extending the two current-carrying tubes of the shunt several diameters beyond the end of the central conductor that serves as a potential lead.

To reduce mutual inductance in the potential circuit of a strip shunt the potential leads can be taken off the strip as a central conductor and coaxial tube as shown in figure 1, b. There still remains the small loop formed by the resistance ribbon itself which is in the potential circuit of the shunt and offers the possibility of inductive pick-up from such currents as might flow in the sheath of the cable and one of the current leads up to the potential circuit. There is no way of separating the potential circuit from end effects arising from such currents in a strip shunt. Although a strip shunt could be built with very low inductive pick-up between its potential circuit and current-carrying parts, the tubular shunt is the only design offering the possibility of entirely eliminating such inductive effects.

In order to understand the importance of being able to ground the sheath of the cable going from the shunt to the CRO at or near the shunt (requirement (3) above), it is necessary to consider the entire surge-current generator circuit. A schematic diagram of a surge-current generator is given in figure 2. The heavy lines show the path of the surge current from capacitor, C, through the tripping gap, test specimen, measuring shunt and the necessary connecting leads. No ground connections are necessary to complete the heavy-current discharge circuit, but this circuit is usually tied to the ground or the floor of the laboratory at some point between the low-voltage terminal of the capacitors and the test specimen in order to complete the charging circuit and to definitely fix the potentials of various parts of the discharge circuit with respect to ground. The self-inductance of the conductors forming the discharge circuit will depend upon their cross-section, length, and arrangement. Assuming the conductors to be of round cross section ½ in. in diameter, their inductance is 0.01 μ h per cm length for a return at a considerable distance. For the maximum rate of change of current 5×10^{10} amp per second, this means the maximum voltage difference between two points 1 cm apart on this discharge circuit is 500 v. Thus, if the discharge circuit were grounded at point g in figure 2, a shunt located only 10 cm from point g would be 5,000 v above ground as the discharge is initiated. This voltage will tend to

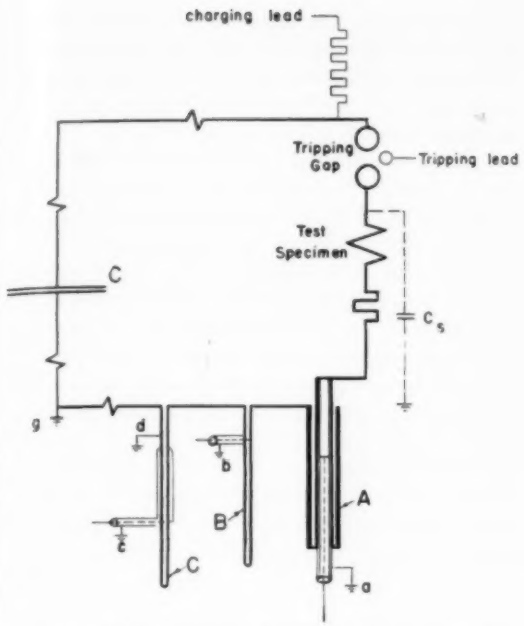


FIGURE 2. *Wiring diagram of surge-current generator showing method of connecting shunts, their potential lead arrangement, and several possible ground connections.*

make current flow in the sheath of the cable from the shunt to the CRO, through the ground connection at the CRO and back to point g through the laboratory grounding system. Due to the stray capacitance (C_s in fig. 2) from the lower ball of the tripping gap and parts of the discharge circuit to ground, this voltage will be of a high frequency oscillatory type occurring just as the main current discharge is initiated and any currents caused by it will induce voltages of the same character in any circuit coupled with them. In order to minimize such inductive effects at the CRO it is necessary that the cable from the shunt to the CRO have its sheath grounded at or near the shunt. In order to reduce the effect of stray capacitance (C_s in fig. 2) it may be desirable to ground other points in the discharge circuit, such as g, in addition to the sheath of the cable. Thus, the best type of shunt would be one which would allow any or all of these ground connections to be made and still have minimum pick-up due to induced voltages from ground currents.

In figure 2, all of the three types of shunts illustrated in figure 1 are represented as being connected in the discharge circuit of a surge generator in order to help explain the effect of ground currents on each. The ground currents most likely to cause inductive pick-up are those due to the high frequency oscillations arising from the stray capacitance, C_s, to ground. If a type C strip shunt were used, a consideration of the various possible ground connections in addition to the cable sheath ground at the CRO, leads to the following conclusions: (1) a single additional ground connection to the cable sheath at c would not be suitable, because then ground currents from C_s would flow in one potential lead of the shunt and induce high voltages in the loop formed by the potential leads; (2) a single additional ground connection at g would still permit cable-sheath currents, which flow to ground at the CRO end of the cable, to flow through one of the shunt potential leads (actually this is better than no ground at g); (3) two additional grounds, one at g and one at c would introduce a combination of the two above effects without eliminating either; (4) a single additional ground at d would reduce the inductive effects of ground currents but would not eliminate them entirely.

Thus the use of a shunt of this type would preclude the experimentally desirable flexibility in location of ground connections.

If a type B strip shunt were used, the same conclusions given for the type C would hold except all inductive effects would be much less because, as may be seen from the diagram, the arrangement of potential leads insures a lower mutual inductance between the potential circuit and current-carrying parts.

If a tubular, type A, shunt were used the ground current for any combination of ground connections would be symmetrically distributed around the axis of the coaxial potential circuit, thus, theoretically at least, eliminating any pick-up due to these ground currents.

The above considerations indicate that the coaxial tubular design is superior to other designs in surge-current measurement work because (1) it should be more constant in impedance over a wide range in frequency, (2) it can be constructed to insure minimum inductive pick-up from current-carrying parts of the surge generator, and (3) it offers the greatest freedom in location of ground connections at the surge generator. In view of these advantages it was decided to investigate

theoretically as well as experimentally the advantages and limitations of concentric tubular shunts for surge current measurements, realizing that some of the derived relations would be applicable to strip-type shunts.

2. Determination of Cross-Sectional Area and Length

The cross-sectional area and length of a shunt are related as follows:

$$V=lA$$
$$R=\rho l/A$$

where V is the volume of the resistance material, l its effective length, A the cross-sectional area, R the resistance, and ρ the volume resistivity.

Solving for A and l

$$A=\sqrt{\frac{V\rho}{R}} \tag{1}$$

$$l=\sqrt{\frac{VR}{\rho}} \tag{2}$$

The resistivity, ρ, is a constant of the shunt material, and it is well to choose a material having as high a resistivity as possible. The resistance of the shunt is determined by the maximum current to be measured and the maximum voltage that can be recorded at the CRO. The volume of the shunt resistance material is fixed by the largest amount of energy that must be absorbed by it during a single discharge of the surge generator. A reasonable temperature rise of say 100° C allows a thermal input of 100 times the specific heat times the mass of the shunt material. For most metals the specific heat is about 0.1 and density is 9, so in general the maximum allowable thermal input to the shunt would be 90 times V, in calories for V in cubic centimeters. The maximum energy fed into the shunt during a single discharge is equal to the energy initially stored in the surge generator capacitors ($1/2\ C\ E^2$), times the resistance of the shunt divided by the minimum effective resistance of the surge-generator circuit. As this minimum effective resistance is usually at least twice that of the shunt, the maximum shunt input energy may be taken as $1/4\ C\ E^2$. When the volume of shunt material computed from this energy input is inserted in the above equations, approximate values of l and A for the shunt are determined. It is best to choose the thinnest wall tubing available having approximately this cross-sectional area as the thinner the wall the slower the change in effective impedance as the frequency is increased (see appendix 1). The limit on wall thinness is set by magnetic stresses, to be discussed later.

If for practical reasons the values actually chosen for l and A differ markedly from the approximate values computed from eq 1 and 2 (this will necessarily be true if the preliminary value of l is too long for convenience or if tubing having approximately the estimated cross-sectional area is not available) the allowable energy input should be computed for the chosen values. This allowable energy input together with the resistance of the shunt and the maximum energy available from the surge generator ($\frac{1}{2}\ CE^2$) determine a minimum effective resistance for the surge-generator discharge circuit. Except in cases where an oscillatory discharge with extremely low damping is required, this minimum effective resistance will be exceeded.

3. Forces on a Tubular Shunt

When a heavy surge current is passed through a tube made of resistance material, it will be subjected to forces from two sources: (1) the action of the magnetic field produced by the current upon the current itself and (2) the thermal expansion of the tube arising from its sudden change in temperature when the current is applied. The magnetic force acts as a pressure from the outside tending to collapse the tube. A formula for computing this pressure is derived in appendix 2. It depends upon the square of the current through the tube and upon its cross-sectional dimensions. Except for extremely high currents most thin-wall tubes available for use in a shunt would probably withstand the pressure arising from the magnetic field; however, in the design of any shunt, and especially in the case of a shunt made with extra-thin-wall tubing, this pressure should be computed in order to make sure that the tube will not collapse.

The sudden change in shunt temperature when the discharge occurs introduces a stress in the tube whose approximate value can be computed. As the shunt tube itself is made of material having much higher resistivity than other current-carrying parts of the shunt and surge generator circuit, it will momentarily attain a higher temperature than its supporting parts during each current discharge. This sudden change in temperature

tends to increase the length of the tube, but it occurs so fast the tube does not have time to expand, and the tube is momentarily put in compression. The maximum compressive stress may be computed from the strain (total temperature change T, times coefficient of linear expansion, δ) and the stress-strain curve for the shunt material. The tube will then behave somewhat like a spring initially compressed and suddenly released; i. e., waves of stress will proceed along the tube, with the velocity of sound, being reflected at the ends and traveling back as tensile stress. The nature and magnitude of these stresses could be computed for any given tube if the exact masses connected to the two ends of the tube were known; however, the maximum stress would not be expected to greatly exceed the initial compressive stress that may easily be computed as noted above. Assuming the shunt material to be perfectly elastic and that it will not be stressed beyond its elastic limit, the maximum stress will be

$$S = ET\delta, \tag{3}$$

where E is the modulus of elasticity for the shunt material. If a value for S is taken equal to the maximum allowable stress (just below yield point) for the material being used, this equation gives an allowable temperature rise for the shunt. For most materials this allowable temperature rise is about 100° C. This value could probably be exceeded because of the high ductility of most resistance alloys; thus if a minimum resistance is desired in the surge generator discharge circuit (maximum energy in the shunt) it would be advisable to determine the maximum allowable temperature rise or energy input experimentally. This could be done by placing a sample of tubing of the same material and cross-sectional dimensions as the shunt tube in the discharge circuit of a surge-current generator. The energy absorbed by the sample may be computed by multiplying the energy initially stored in the capacitors ($\frac{1}{2} CE^2$) by the ratio of the resistance of the sample to the total effective resistance of the complete surge generator discharge circuit including the sample tube. The total effective resistance of the surge generator can be deduced by using a shunt, whose resistance tube has a cross-sectional area at least five times that of the sample tube, in the discharge circuit and taking oscillograms of the discharge current. From measurements of frequency and logarithmic decrement on these oscillograms, the inductance and effective resistance of the complete surge-generator discharge circuit can be computed. Thus, by starting with the surge generator capacitors charged to a fairly low voltage, E, and taking repeated discharges through the sample tube at successively higher values of E the energy input at which the tube is broken or deformed may be determined. The safe allowable temperature rise for a tube of this particular material and size can then be stated.

The experimental procedure just described was carried out for a $\frac{5}{16}$-in. diameter CuNi alloy tube whose wall thickness was 0.039 cm and length 33 cm. No evidence of tube failure except discoloration due to the high temperature was noted for energy inputs below 160 joules/centimeter (instantaneous temperature rise of 400° C). However, after the first discharge at this energy input a slight bend in the tube was noted as though it were failing as a column in compression. After each discharge of successively higher energy input the bending was more pronounced and after a discharge at 268 j/cm (670° C temperature rise) in addition to the bending, decided crinkling and crushing were noted. The crushing was probably caused by magnetic forces acting at an instant when the tube was weakened because of its high temperature. These values of energy input exceed what might be considered a desirable design value.

The safe design value should be one that would not stress the tube beyond its elastic limit. This was determined in another experiment by measuring the length of the tube before and after each discharge. After two discharges, the energy input of each being equal to 60 j/cm (150° C temperature rise), no permanent change in length was noted. For energy inputs of 100 j/cm (260° C temperature rise) or greater an increase in length of about 0.2 mm (0.6 percent) was noted after each discharge. Thus, it was concluded that the maximum allowable temperature rise for a shunt made of this material should be about 150° C. This is only slightly above the maximum value (133° C) computed from the stress due to thermal expansion.

4. Inductance and Capacitance Between Current Terminals

Because of the method of attaching potential leads, the 4-terminal impedance of a tubular shunt

is constant over a wide range of frequency. However, at very high rates of change of current (high-frequency components of the surge), the distributed capacitance and inductance of the current circuit will cause variations in the root-mean-square values of these high-frequency components along the length of the shunt. That is, the instantaneous value of current along the length may differ from the value at the current terminals.

This variation in the instantaneous value of current along the length of the shunt is, of course, entirely negligible at lower frequencies, but in order to determine the frequency at which this effect becomes appreciable a solution for the current and voltage along the length of the tube, assuming uniformly distributed inductance and capacitance only, was obtained (in a manner similar to that used in appendices 3 and 4). The current at the potential lead end of the shunt, I_l, was found to be

$$I_l = \frac{I_c}{\cos \omega l \sqrt{LC}} \tag{4}$$

where

I_c=current at current terminals; L=distributed inductance per unit length of shunt; C=distributed capacitance per unit length of shunt; $\omega = 2\pi f$ (f in cycles per second); l=length of current circuit in centimeters.

For concentric tubes whose wall thickness is small compared to their diameter, the inductance and capacitance per unit length are:

$$L = 2 \log_e b/a \times 10^{-9} \text{ henries} \tag{5}$$

and

$$C = \frac{0.555}{\log_e b/a} \times 10^{-12} \text{ farads} \tag{6}$$

where a is the outer diameter of the inner tube, and b is the inner diameter of the outer tube. Multiplying eq 5 by eq 6 gives $LC = 1.110 \times 10^{-21}$. Thus the quantity $\sqrt{LC}$ in eq 4 is a constant, and the variation in current along the length of the tubes depends only upon the length, l, of the tubes and the frequency of the current being measured. For shunts up to 100 cm long, the maximum variation in current is less than 10 percent for frequencies up to 21.5 megacycles.

The voltage between the current terminals was found to be

$$E_3 = j\omega LlI\left(1 + \frac{\omega^2 LC}{3} l^2 + \ldots\right) \tag{7}$$

The magnitude of this voltage does not affect the accuracy of the shunt, but it should be evaluated to indicate the insulation required between current terminals. For shunts up to 100 cm long the quantity in the bracket of eq 7 reduces to unity (within 5 percent) for frequencies up to 18 megacycles. Thus this voltage may be taken as equal to $j\omega LlI$ or $LldI/dt$. The total inductance of the current circuit for any tubular shunt, Ll, may be computed from its dimensions by use of eq 5, and the rate of change of current dI/dt depends upon the particular surge generator being used and its discharge circuit. For a typical shunt, 100 cm long, this inductance is about 0.14×10^{-6} henries, and for the maximum rate of change of current assumed above (5×10^{10} amp/sec) the voltage between current terminals is 7,000 v.

III. Description of Two Shunts

There are two surge generators being used in this laboratory at the present time. One is a 2,000 kv surge-voltage generator capable of giving 30,000 amp on short circuit. The other is a 100 kv 10μf surge-current generator capable of discharging a maximum of 200,000 amp. It was

TABLE 1.—*Constants of two tubular shunts*

	Shunt A	Shunt B
Four-terminal resistance	0.04 ohm	0.0048 ohm
Over-all length	39 in	38¼ in.
Length of resistance tube	26⅞ in	24⅞ in.
Diameter of outside tube	⅞ in	2 in.
Diameter of resistance tube	¼ in	1 in.
Wall thickness of resistance tube	0.008 in. (0.0203 cm).	0.0246 in. (0.0 cm).
Cross-sectional area of resistance tube	0.0392 sq cm	0.469 sq cm
Resistance tube material	German silver	80% Cu, 20% Ni
Resistivity of resistance tube (micro-ohm cm).	23.33	36.3
Ratio of a-c impedance to d-c resistance, at:		
f=0.25 megacycle	1.00	0.98
f=0.50 megacycle	1.00	0.92
f=1.0 megacycles	0.99	0.76
f=2.0 megacycles	0.96	0.48
f=5.0 megacycles	0.86	0.15
Allowable temperature rise—computed from stress due to thermal expansion	133°C	133°C
Allowable energy input	1320 joules	14,500 joules
Current limit due to magnetic force	87,600 amp	223,000 amp
Total inductance of current circuit	0.202 μ h	0.100 μ h
Maximum voltage drop at current terminals—$L\,di/dt$ for $di/dt = 5 \times 10^{10}$ amp/sec	10,000 v	5,300 v

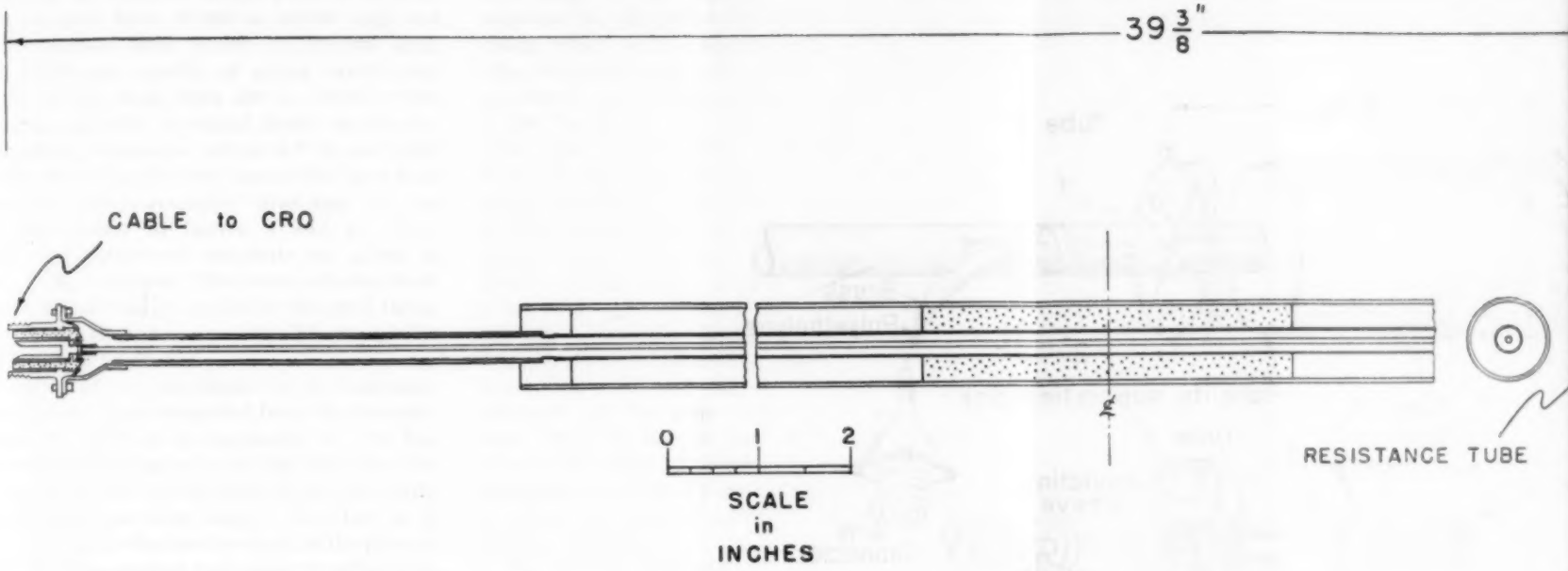
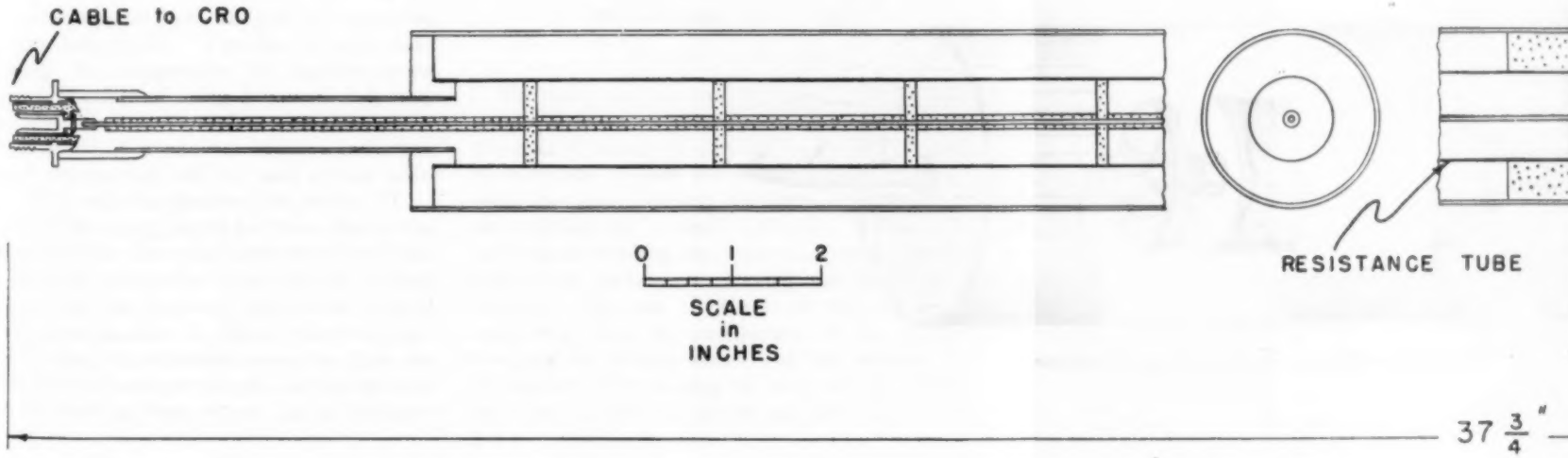


FIGURE 3.—*Structural details of small tubular shunt, A.*
See table 1 for additional dimensions and other constants.

FIGURE 4.—*Structural details of large tubular shunt, B.*
See table 1 for additional dimensions and other constants.



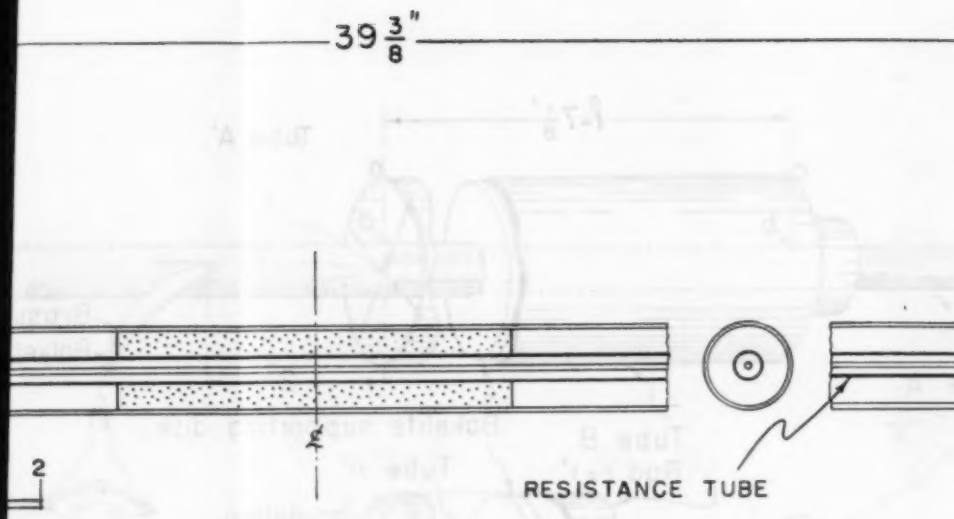
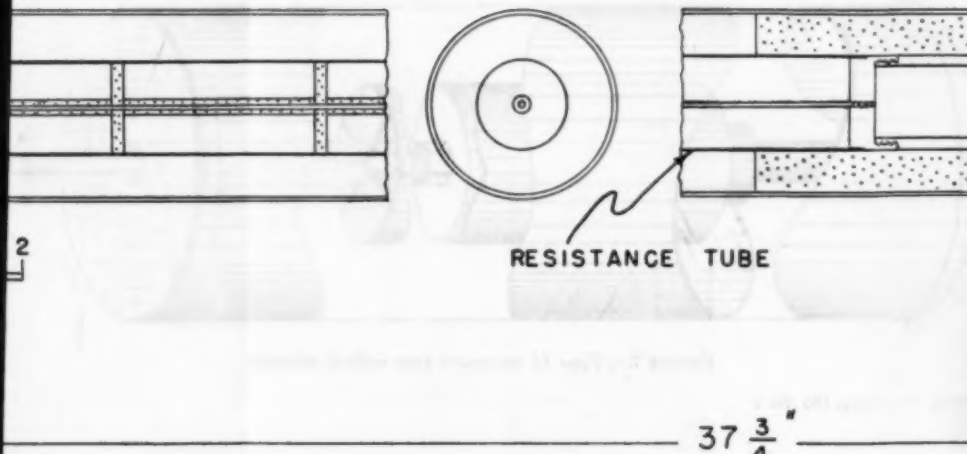
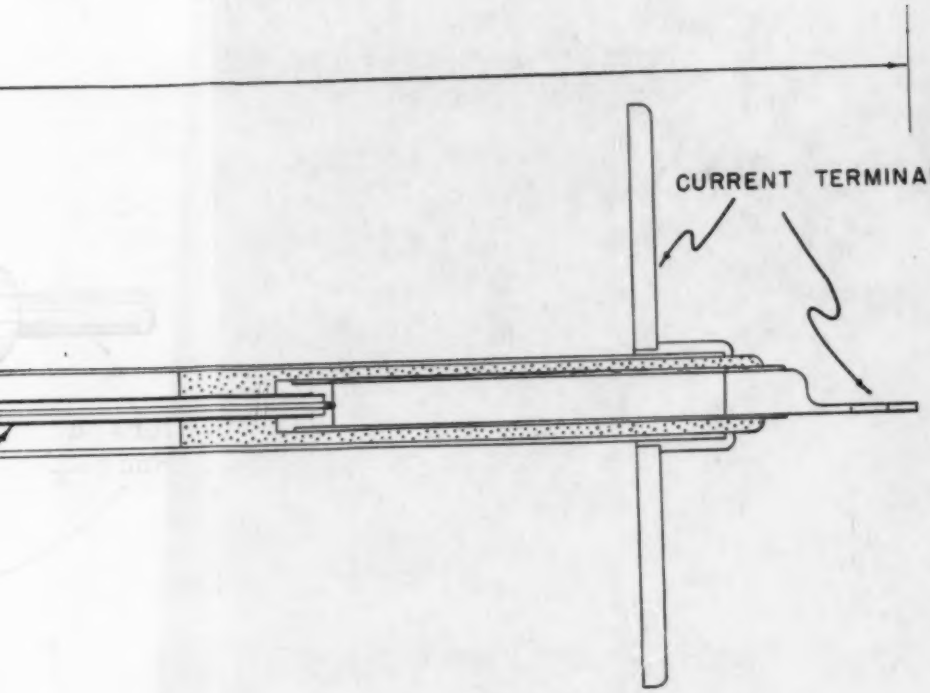
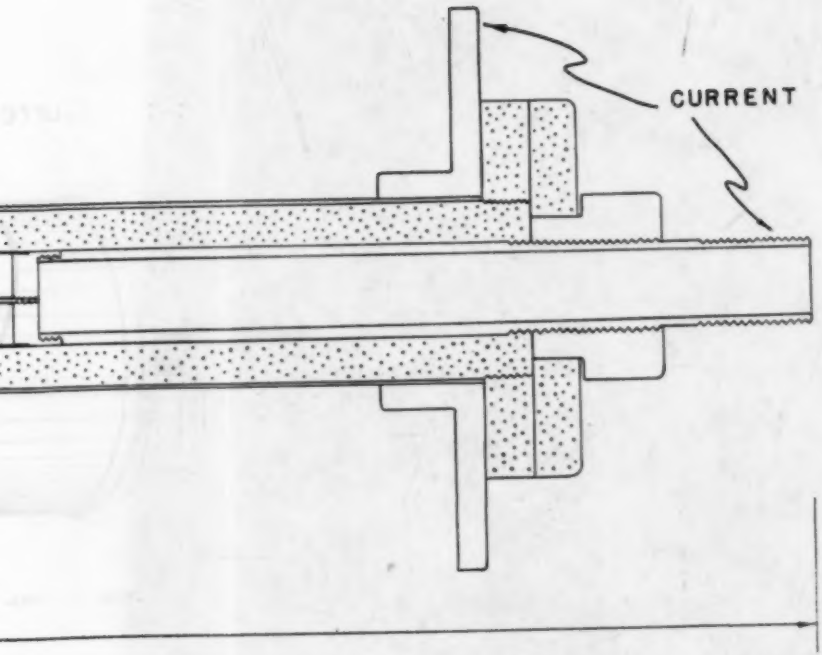


FIGURE 3.—*Structural details of small tubular shunt, A.*
See table 1 for additional dimensions and other constants.

FIGURE 4.—*Structural details of large tubular shunt, B.*
See table 1 for additional dimensions and other constants.







UNIV. OF MICH.

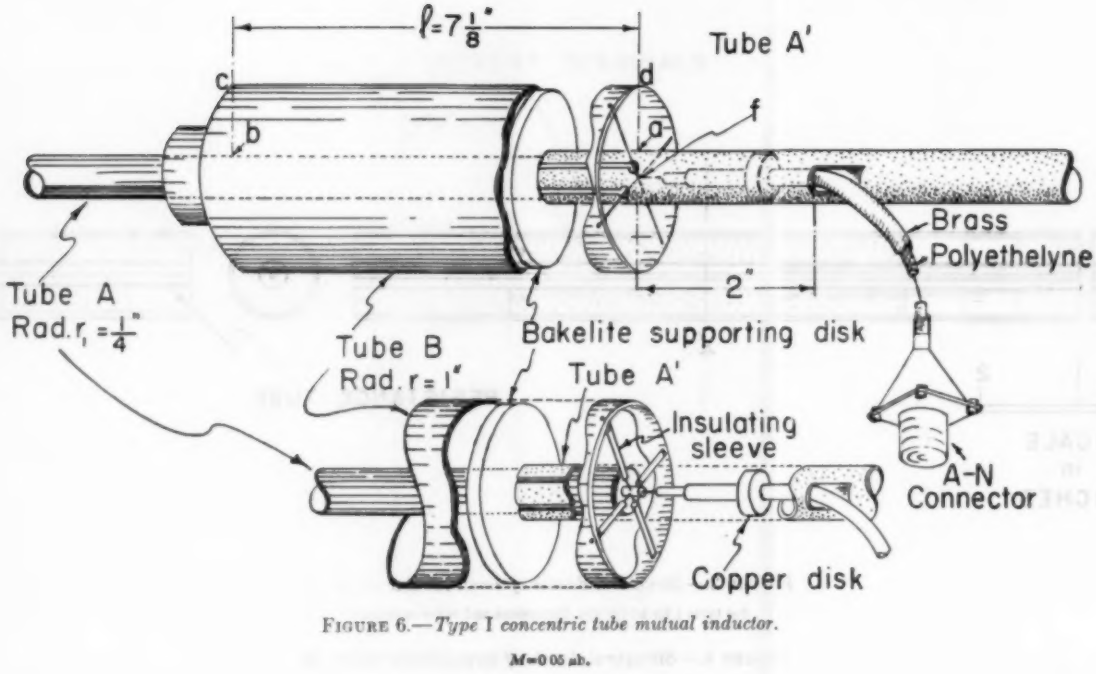


FIGURE 6.—*Type I concentric tube mutual inductor.*

$M = 0.05$ mh.

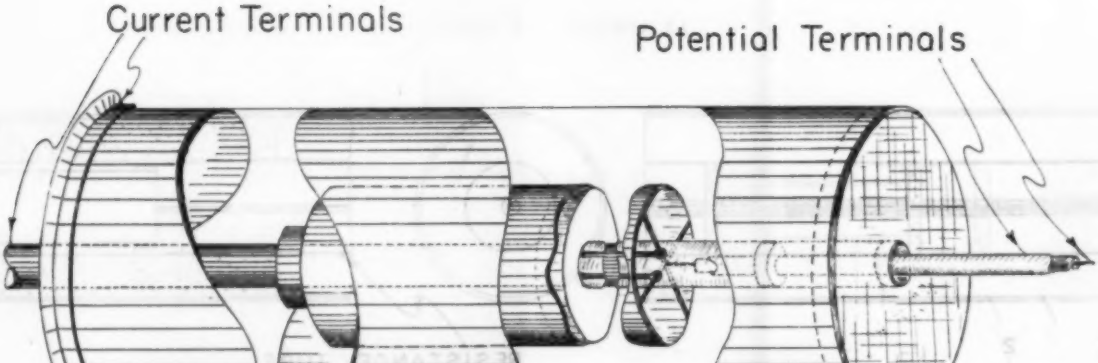


FIGURE 7.—*Type II concentric tube mutual inductor.*

thought that at least two shunts would be needed, one for currents from 10,000 to 50,000 amp and one for currents from 50,000 to 200,000 amp. As the CRO was capable of giving satisfactory records over the range from 200 to 2,000 v when used with a suitable potential divider at the deflecting plates, resistance values of 0.04 and 0.005 ohm were chosen for the two shunts that have been constructed. Cross-sectional drawings of the shunts are shown in figures 3 and 4. Their physical and computed constants are given in table 1. The resistance tubes were selected from a small stock of readily available thin-wall tubes, because at the time it was not feasible to wait for special tubes to be drawn.

The ratios of *a-c* impedance to *d-c* resistance given in table 1 were computed from the theoretical formulae derived in appendix 1. At frequencies above 5 megacycles for the 0.04 ohm and 1 megacycle for the 0.0048 ohm shunt, this ratio is considerably less than unity. This fact is of no great concern in the measurement of the present standard impulse-current test waves for which the fundamental-frequency component is of the order of 100 kc or less; but the fact should not be overlooked that the magnitude of higher-frequency components, probably present in such test waves, may be considerably reduced on a comparative basis. Other types of shunts, as they have some inductive coupling between their potential-lead circuit and current-carrying parts of the discharge circuit, may give an indication of these higher frequency superposed currents. However, this indication will depend, not upon the resistance of the shunt, but upon the unknown value of inductive coupling, and thus will give an erroneous idea of their magnitude. A method of accurately determining the magnitude of high-frequency components consists of using a mutual inductor, having a known value of coupling with the current circuit, in place of a shunt. Mutual inductors have been constructed and are used at this laboratory. They will be discussed in section IV.

The allowable energy input for these shunts was computed from the allowable temperature rise (obtained from the maximum stress due to thermal expansion) and the mass of each shunt (actual values are given in table 1). These values are useful in computing the minimum resistance that can be used in the discharge circuit of a surge generator without overheating these shunts. As an example: The surge voltage generator in this laboratory requires the dissipation of 33,300 j for each discharge when charged to full-rated voltage. Thus, when shunt A is used with this generator, the total resistance of the discharge circuit must be at least 33,300 divided by 1,320 times the shunt resistance (0.04 ohm) or 1.01 ohms. A similar computation for shunt B gives 0.107 ohm. As the minimum discharge resistance of this surge voltage generator is 6 ohms, either shunt may be used with any discharge circuit. The surge-current generator in this laboratory requires the dissipation of 50,000 joules when charged to full-rated voltage, thus the total resistance of the discharge circuit must be at least 1.52 ohms if shunt A is to be used or 0.0165 ohm if shunt B is to be used. As the minimum discharge resistance of this surge-current generator is 0.03 ohm, shunt B may be used with any discharge circuit, but shunt A should only be used when a resistance of at least 1.5 ohms is added (a resistance of about this value is needed to give a critically damped discharge).

The current limitations on these shunts due to magnetic forces were estimated by the formula given in appendix 2. These limitations are above the values of current expected to be used with these shunts. The inductances of the current circuits of these shunts were computed by using the formula given in eq. 5. The voltage drops computed from these inductances and the maximum expected rate of change of current indicate the desirability of having adequate insulation between the current terminals to prevent flashover.

IV. Mutual Inductors for Measuring Rates of Change of Current

1. General Considerations

For many experiments involving heavy-current surges, it is desirable to know the wave form of the rate of change of current, for example: (1) in studying the voltage induced in a circuit coupled with the heavy current discharge circuit and (2) in studying the voltage difference arising from inductance between two points on a conductor or a system of conductors carrying the heavy-surge current. The rate of change of current can be computed from an oscillogram of the current obtained by using a shunt, but this involves the inaccuracy of measuring the slope of the current wave at a number of points and plotting a curve

from these measured values of slope. In addition, the high-frequency components of the current wave form as they appear in the oscillogram obtained with a shunt may be highly distorted or attenuated and give an erroneous value of rate of change of current, especially at the point of maximum rate of change. A more accurate record of rate of change of current can be obtained with much less effort by inserting the primary of a mutual inductor in the heavy-current discharge circuit in place of the shunt and connecting the secondary of this inductor to the cable to the CRO. The record obtained on the CRO is then equal to $M\ di/dt$. If M (the mutual inductance of the inductor) is known, an actual oscillogram of the rate of change of current is obtained.

A mutual inductor, to be considered satisfactory for this purpose, must fulfill the following requirements: (1) Its mutual inductance must be of suitable value to give the proper voltage at the CRO (from 500 to 2,000 in this case). For the expected rates of change of current (from 1 to 5 times 10^{10} amp per second) this amounts to about 0.05 μh. This value of mutual inductance must be definite and should be computable from the physical dimensions of the inductor.

(2) The primary circuit of the inductor must be capable of carrying the heavy current discharge and should add a minimum impedance to the discharge circuit.

(3) The secondary circuit should have minimum coupling to all current carrying parts except the primary of the inductor.

(4) The method of connecting the secondary circuit to the cable going to the CRO should be such that the sheath of this cable can be connected to ground near the surge current generator without introducing coupling between ground currents in the sheath and the secondary circuit.

(5) The secondary circuit should have a minimum self-inductance. This may be explained as follows: The voltage to be measured at the CRO ($M\ dI/dt$) is impressed upon a circuit containing the self-inductance, L, of the secondary circuit and the surge impedance, R, of the cable to the CRO. It is the voltage across the surge impedance, R, that is recorded at the CRO and its value is

$$E_{\text{CRO}}=\frac{M\ dI/dt}{1+j\frac{\omega L}{R}}. \tag{8}$$

In order to have E_{CRO} equal to $M\ dI/dt$ for the high frequency components of I, it is apparent that the self-inductance, L, of the mutual inductor secondary circuit should be kept as low as possible. As L cannot be reduced to zero, the highest frequency components of I are subject to a phase displacement and attenuation, the magnitudes of which can be computed from the above equation for any assumed value of $\omega=(2\pi f)$.

A consideration of the various convention types of mutual inductors leads to the conclusion that none of them will fulfill the above requirements satisfactorily. The most promising type appeared to be one consisting of a straight single conductor primary insulated from, and running through, the central axis of a toroidal coil secondary of sufficient turns to provide adequate voltage at the CRO. It was thought that by introducing shields and connecting them to the primary circuit at an appropriate point with respect to the ground connection, an inductor of this type might fulfill most of the requirements listed above. A preliminary design of an inductor of this type was carried out, but the self-inductance of its secondary circuit had to be at least 2 μh in order to give the desired value of mutual inductance. When this value of inductance and 50 ohms for the CRO cable surge impedance, R, are inserted in eq. 8, it is found that the voltage at the CRO is only 0.9 of $M\frac{dI}{dt}$ for a frequency of 1.8 megacycles, for higher frequencies the attenuation would be much greater. The use of a mutual inductor of this type was, therefore, abandoned.

2. Mutual Inductors of Special Design

The first attempt to design a special mutual inductor to meet all of the requirements listed above resulted in the inductor shown in figure 5. The primary circuit is the straight length of No. 10 copper wire (with the return circuit at least 2 ft away, its effect can be neglected). The secondary circuit consists of the "D" formed by part of the No. 10 wire and the No. 16 wire. This arrangement permits the magnetic flux linking the secondary "D" to be computed as that due to an infinitely long, straight conductor, neglecting the effect of the return lead. The self-inductance of the secondary circuit is that due to the single "D" loop and can also be computed from its dimensions.

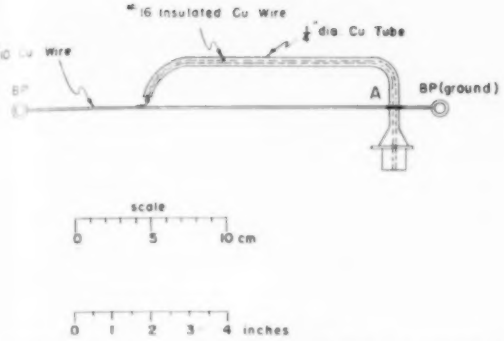


FIGURE 5.—*D-type mutual inductor used to obtain records of rate of change of surge currents.*

M=0.0986 μh

An inductor of this design, with dimensions corresponding to the scale in figure 5 was constructed. Its mutual inductance was computed to be 0.0986 μh and the self-inductance of its secondary circuit 0.25 μh. This value of self-inductance is considerably less than that of the toroidal secondary circuit mentioned earlier, and as a consequence frequencies up to 15 megacycles will be recorded with only 10 percent reduction in magnitude. The secondary circuit is connected to the CRO cable by coaxial leads, thus no voltage will be induced in it from ground currents flowing in the cable sheath. A mutual inductor of this type appears to fulfill the requirements of surge current testing fairly well. The one constructed in this laboratory has been of considerable value in obtaining records of the rate of change of surge currents. Its outstanding advantage lies in simplicity of construction.

The "D"-type mutual inductor has some drawbacks: (1) it is not completely unaffected by magnetic fields produced by other current-carrying parts of the surge generator circuit and (2) the self-inductance of its secondary circuit is about three times its mutual inductance (as a minimum it should be possible to make an inductor whose self-inductance is as low as its mutual inductance). Another mutual inductor has been designed in which the secondary circuit consists of a tube arranged coaxially with the primary conductor in order to overcome the small defects inherent in the "D"-type inductor.

A type I coaxial tube mutual inductor is shown in figure 6. Tube A constitutes the primary circuit of this inductor (the surge-current return lead is again assumed to be located at least 24 in. away so its effect can be neglected). The secondary circuit of this inductor consists of the following parts: (1) tube, A, from a to b, (2) a radial ring soldered on its inside to and around tube A at b and soldered on its outside to and around the inside of tube, B, at c, (3) tube B from c to d, and (4) the four-wire radial "spider" connection from d on tube B, going through but insulated from tube, A, and soldered to the central conductor of the potential leads at f. The secondary terminals are thus at a and f and are brought out through the concentric leads to the fitting for the CRO cable. How well this design of mutual inductor fulfills each of the requirements listed above may best be seen by considering each requirement, as follows: (1) The mutual inductance can be computed as that due to the magnetic flux produced by current in a long straight conductor (tube A) integrated over the region $a\ b\ c\ d\ a$ between the tubes. The mutual inductance thus derived in terms of the tube dimensions indicated on figure 6 is

$$M=2l\log_e\frac{r_2}{r_1}10^{-9}\text{ henry (for } l \text{ in cm)}.$$

There is a small correction to this value of inductance due to the thickness of tubes A and B, which can be computed by the method explained on page 397 of reference [4]. For $r_1=0.25$ in. and $r_2=1$ in., the length, l, of tube B should be about 7 in. to give a mutual inductance of 0.05 μh. (2) Because of the method of bringing out the secondary circuit leads used in this design, tube A must be at least 0.5 in. in diameter, and it should be made of a good conducting material such as copper. The wall thickness of this tube is relatively unimportant, and any ordinary thin wall copper tube can be used that is adequate to carry the heavy current discharge without damage. (3) The secondary circuit except for the ring and spider are coaxial conductors and thus have minimum coupling to all currents not flowing in the inductor. Likewise the radial ring and spider conductors are symmetrical about the axis which results in minimum coupling. (4) Also, because of the coaxial arrangement of the secondary circuit and leads to the CRO cable, any ground currents flowing in the cable sheath will have minimum

coupling with the secondary circuit and leads. (5) The self-inductance of the secondary circuit can be computed by the method explained on page 397 of reference [4], and it is found to be equal to the mutual inductance except for a small term arising from the finite thickness of tube, B. The self-inductance of the spider alone is less than 0.005 μh. Thus this design gives the minimum self-inductance that can be obtained in a straightforward manner.

An approximate value of the voltage impressed on the CRO cable can be obtained from eq. 8, but this equation neglects the capacitance between tube, A, and tube, B, and will not hold for the higher frequencies. A more accurate method for computing the value of this voltage is to consider the capacitance and inductance of the secondary circuit to be uniformly distributed along its length and solve for the voltage at the CRO end of this circuit by the methods normally used for circuits of uniformly distributed constants. This is done in appendix 3, and as shown there for a typical mutual inductor designed for surge-current measurements, the voltage impressed on the CRO cable will be equal to $M\ dI/dt$ (within 10 percent) for all frequencies up to 65 megacycles/sec.

The type I tubular mutual inductor appears to satisfy the requirements for surge-current measurement quite well. One was constructed in this laboratory having the approximate dimensions indicated in figure 6. This type of inductor is entirely satisfactory if the surge-generator discharge circuit is fairly long and can be arranged so that the inductor forms the final part of the return lead to ground. In case the minimum length of discharge circuit is required, the length of this inductor may add an appreciable inductance to the discharge circuit and it may also not be feasible to keep the return lead 24 in. away from the axis of the inductor. For such cases, a slightly different design of tubular inductor (type II) as indicated in figure 7 is suggested. It is the same as the type I inductor except that a return circuit is provided for the current by a larger tube, coaxial with the smaller tubes. This design brings the two current terminals close together and eliminates the problem of the location of the inductor with respect to other parts of the surge-generator circuit. It allows the inductor to be used in a discharge circuit of minimum length with the addition of the least possible inductance to the discharge circuit. The outer current return tube of the type II inductor also acts as an electrostatic shield for its secondary circuit. Because of the distributed capacitance between the outer-current return tube and the coaxial secondary circuit tube, currents and voltages in the secondary circuit will not be the same for the type II inductor as for the type I. Thus a computation of the voltage impressed on the CRO cable for a type II inductor must take into account the distributed capacitance and inductance of the circuit formed by the outer tube of the secondary circuit and the return current tube in addition to that of the secondary circuit itself. This is done in appendix 4, and while the results appear quite different from those obtained for the type I inductor in appendix 3, they indicate about the same upper limit in frequency (70 megacycles/sec).

Each of the three types of mutual inductors described above should prove useful for surge-current measurements. The "D" mutual inductor is very simple to construct and will measure superposed frequencies up to 15 megacycles/sec with less than 10 percent error. The tubular inductors are somewhat more complicated to construct but extend the frequency range to 70 megacycles/sec. The type I tubular inductor is suitable when the surge-generator discharge path is fairly long and the type II when a minimum length of discharge path is required. Inductors of both the "D"-type and the type I tubular have been constructed and proved very useful in this laboratory.

V. Experimental Results

The shunts and mutual inductors described above have been used in this laboratory for measuring current surges both from a surge-voltage generator and a surge-current generator and for various arrangements of discharge circuits. The oscillograms in figure 8 are typical of the results obtained. They are records of the discharge current and rate of change of current for a surge-current generator consisting of 40 1-μf, 50-kv, capacitors connected in series-parallel so as to give a total capacitance of 10 μf and a voltage rating of 100 kv. The discharge circuit consisted of a

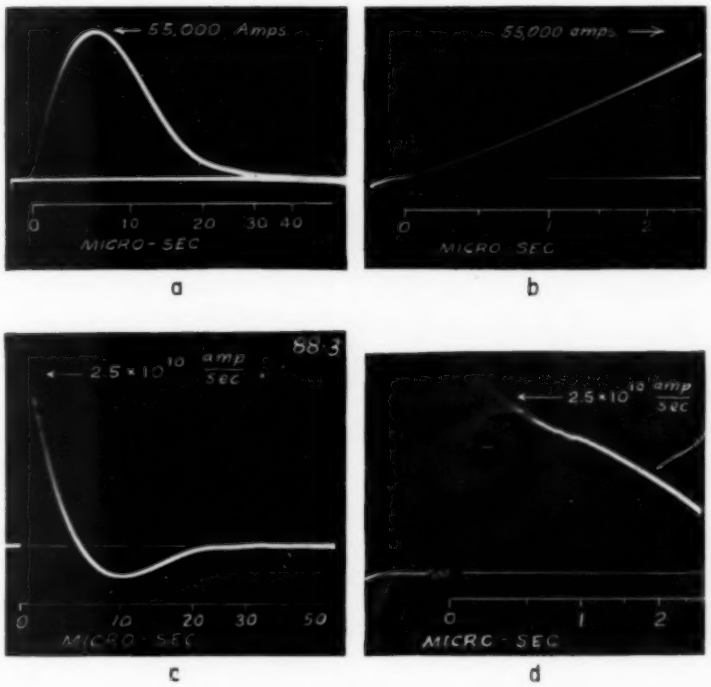


FIGURE 8.—CRO *records of a critically damped discharge from a surge-current generator.*

a, Current wave-form obtained by use of tubular shunt, *A*. *b*, early part of current wave-form on a faster CRO sweep. *c*, rate of change of current obtained by use of the mutual inductor shown in figure 6; *d*, early part of rate of change on a faster CRO sweep.

three-ball triggering gap; a 1-ohm resistor made up of 24 globar type B ceramic resistor units, assembled in the form of a cage (to minimize inductance) having sets of two 12-in. units in series, connected in parallel so that this assembly was 24 in. long; about 5 ft of copper busbar; and the shunt or mutual inductor. The total inductance of this generator and its discharge circuit is 2.5 μh and the 1-ohm resistor gives very close to critical damping. For the oscillograms shown in figure 8 the generator was charged to 75 kv and gave a maximum discharge current of 55,000 amp, the maximum rate of change of current being 2.5×10^{10} amp/sec

The oscillogram in figure 8, a, shows the variation of current with time for the entire surge. Figure 8, b, shows the initial rise of the current on an expanded time scale (faster CRO sweep). Figure 8, c, shows the variation in the rate of change of current for the entire surge, and figure 8, d, shows the initial rate of change of current using a faster sweep. The current records are quite smooth and just as anticipated from theory, as are the rate of change of current records except for the superposed high-frequency oscillations occurring just as the surge is initiated. These superposed oscillations have a frequency of about 40 megacycles and are probably caused by stray capacitance across parts of the discharge circuit, such as from one ball of the tripping gap to ground. If this stray capacitance is across a section of the discharge circuit having an inductance of 2 μh, it would require only about 10 $\mu\mu$f to produce the 40 megacycle oscillations. It is practically impossible to eliminate such stray capacitances. In the discharge circuit used in obtaining the oscillograms of figure 8, stray capacitances were reduced to as low a value as possible. For discharge circuits involving fairly large apparatus as a test specimen, stray capacitances will be larger, resulting in a superposed oscillation of lower frequency.

The slight upward curvature on the zero line

of oscillograms in figures 8, b and 8, d at the start of the sweep is introduced by the oscillograph and should not be considered as part of the surge record. The source of this "hook" in the CRO zero line is a small horizontal deflection given to the beam by the Norinder relay plates as voltage is initially applied to them. This deflection slightly alters the CRO sweep rate at its beginning but should, by itself, introduce no vertical deflection. However, since the focusing coil is located below the Norinder relay plates, this deflection puts the beam slightly off the axis of the focusing coil which then produces the small vertical deflection.

The oscillograms in figure 8 illustrate the advantages of using a mutual inductor in addition to a shunt when measurements of a current surge are being made. The superposed oscillations cannot be detected on the current record but can readily be seen and measured on the rate-of-change-of-current record. Their actual magnitudes in amperes can be computed from this record, and from measurements on the oscillogram of figure 8, d they were found to have a maximum amplitude of about 25 amp. Oscillations of this magnitude would be almost impossible to detect on a current record even though the attenuation of the shunt for these higher frequencies did not mask them out. The rate of change of current records are also of great value in studying the effects of induced voltages arising from the current surge. They further provide an accurate measurement for the time to maximum current (zero rate of change).

VI. References

[1] P. L. Bellaschi, Heavy surge currents—generation and measurement, Electrical Eng. **53,** 86 (Jan. 1934).

[2] T. Brownlee, Discussion of Bellaschi's paper, Heavy surge currents—generation and measurement, Electrical Eng. **53,** 481 (Mar. 1934).

[3] N. Rohats, A surge-current generator, General Electric Rev. **37,** 296 (June 1934).

[4] F. B. Silsbee, A study of the inductance of four-terminal resistance standards, Bul BS **13,** 375 (1916) S281.

[5] F. B. Silsbee, Notes on the design of four-terminal resistance standards for alternating current, BS J. Research **4,** 73 (1930). RP133.

VII. Appendices

1. Effective Impedance of a Tubular Shunt as a Function of Frequency [6]

Let the radii of the inner and outer tubes be a, b, a', and b' as indicated in figure 9. Assuming the tubes to be coaxial and perfectly symmetrical about their axis the current density in either tube at a given radius, r, from the axis will be a constant for any given instant, so let

i_r = current density in inner tube at radius r, + if into paper

i_r' = current density in outer tube at radius r, + if into paper

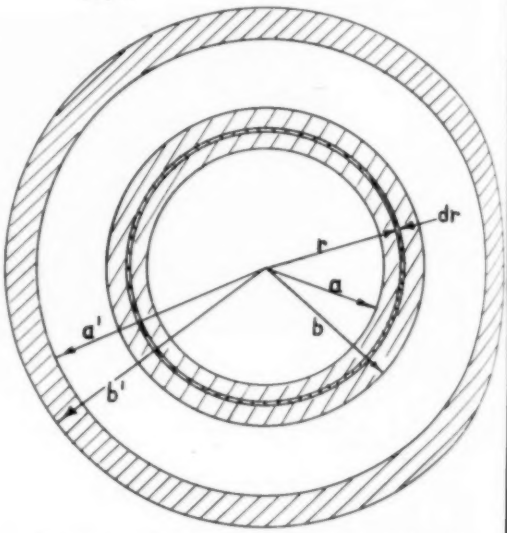


FIGURE 9.—*Cross-sectional view of a tubular shunt.*

Then

I = total instaneous current in inner tube
$= \int_b^a 2\pi i_r dr$

I' = total instantaneous current in outer tube
$= \int_{b'}^{a'} 2\pi i_r' r dr$

Also, let:

e = emf drop per unit length of inner tube
e' = emf drop per unit length of outer tube
H_r = magnetic field at r taken + if clockwise
ρ = resistivity of inner tube material
ρ' = resistivity of outer tube material
Φ_r = total magnetic flux outside of radius r (all permeabilities assumed = 1).

NOTE.—The cgs magnetic system of units is used throughout this derivation and all dimensions are in centimeters.

[6] This derivation is taken from hitherto unpublished notes of F. B. Silsbee.

As all currents flow along lines parallel to the tube axis, e and e' are independent of r, so

$$e=\rho i_r+\frac{\partial \Phi_r}{\partial t} \tag{9}$$

$$e'=\rho' i_r'+\frac{\partial \Phi_r'}{\partial t} \tag{10}$$

and

$$\frac{\partial e}{\partial r}=\rho\frac{\partial i_r}{\partial r}+\frac{\partial}{\partial t}\frac{\partial \Phi_r}{\partial r}=0 \tag{11}$$

$$\frac{\partial e'}{\partial r}=\rho'\frac{\partial i_r'}{\partial r}+\frac{\partial}{\partial t}\frac{\partial \Phi_r'}{\partial r}=0 \tag{12}$$

For the inner tube

$$\Phi_r=\int_r^b \frac{2I_x}{x}\,dx+2I\log\frac{a'}{b}+\int_{a'}^{b'}\frac{2(I+I_x')}{x}\,dx \tag{13}$$

where

$$I_x=\int_a^x 2\pi i_r rdr \text{ and } I_x'=\int_{a'}^x 2\pi i_r' rdr$$

For $r>b'$, $\Phi_r=0$. Differentiating eq 13 with respect to r

$$\frac{\partial \Phi_r}{\partial r}=-\frac{2I_r}{r}=-\frac{2}{r}\int_a^r 2\pi i_r rdr \tag{14}$$

Similarly for the outer tube,

$$\Phi_r'=\int_r^{b'}\frac{2(I+I_x')}{x}\,dx \tag{15}$$

and

$$\frac{\partial \Phi_r'}{\partial r}=-\frac{2(I+I_x')}{r}=-\frac{2}{r}\left[I+\int_{a'}^r 2\pi i_r' rdr\right]=+\frac{2}{r}\int_r^{b'} 2\pi i_r' rdr \tag{16}$$

since $I=-\int_{a'}^{b'} 2\pi i_r' rdr$

Inserting eq 14 in eq 11 and eq 16 in eq 12,

$$r\frac{\partial i_r}{\partial r}=\frac{4\pi}{\rho}\frac{\partial}{\partial t}\int_a^r i_r rdr \tag{17}$$

$$r\frac{\partial i_r'}{\partial r}=-\frac{4\pi}{\rho'}\frac{\partial}{\partial t}\int_r^b i_r' rdr \tag{18}$$

Differentiating again with respect to r

$$r\frac{\partial^2 i_r}{\partial r^2}+\frac{\partial i_r}{\partial r}-\frac{4\pi}{\rho}r\frac{\partial i_r}{\partial t}=0 \tag{19}$$

$$r\frac{\partial^2 i_r'}{\partial r^2}+\frac{\partial i_r'}{\partial r}-\frac{4\pi}{\rho'}r\frac{\partial i_r'}{\partial t}=0 \tag{20}$$

If $i_r=I_0e^{j\omega t}$

$$\frac{\partial}{\partial t}i_r=+j\omega i_r$$

and eq 19 and 20 become

$$\frac{\partial^2 i_r}{\partial r^2}+\frac{1}{r}\frac{\partial i_r}{\partial r}+K^2 i_r=0 \tag{21}$$

$$\frac{\partial^2 i_r'}{\partial r^2}+\frac{1}{r}\frac{\partial i_r'}{\partial r}+K'^2 i_r'=0 \tag{22}$$

where $K^2=-j\omega\frac{4\pi}{\rho}$ and $K'^2=-j\omega\frac{4\pi}{\rho'}$, the solution of eq 21 is

$$i_r=Z_0(Kr)=C_1J_0(Kr)+C_2N_0(Kr) \tag{23}$$

where J_0 and N_0 are Bessel's functions of order 0 and of the first and second kind. C_1 and C_2 are constants of integration. Similarly the solution to eq 22 is

$$i_r'=Z_0'(K'r)=C_1'J_0(K'r)+C_2'N_0'(K'r) \tag{24}$$

The total current is

$$I=2\pi\int_a^b i_r rdr=2\pi\int_a^b rZ_0(Kr)dr \tag{25}$$

and from the general relation

$$\int xZ_0(ax)dx=\frac{1}{a}xZ_1(ax) \tag{26}$$

it follows that

$$I=\frac{2\pi}{K}[bZ_1(Kb)-aZ_1(Ka)] \tag{27}$$

Similarly

$$I'=\frac{2\pi}{K}[b'Z_1'(K'b')-a'Z_1'(K'a')] \tag{28}$$

Likewise

$$I_x=\frac{2\pi}{K}[xZ_1(Kx)-aZ_1(Ka)] \tag{29}$$

$$I_x'=\frac{2\pi}{K'}\left[xZ_1'(K'x)-a'Z_1'(K'a')\right] \tag{30}$$

755528—47——2

To obtain values for the Φ_r's substitute eq 29 and 30 in eq 13 and 15. For the inner tube, after performing the integrations

$$\Phi_r=\frac{4\pi}{K^2}\left[-Z_0(Kx)\right]_r^b-\frac{4\pi aZ_1(Ka)}{K}\log\frac{b}{r}+2I\log\frac{b'}{b}+\frac{4\pi}{K'^2}\left[-Z_0'(K'x)\right]_a^{b'}-\frac{4\pi a'Z_1'(K'a')}{K'}\log\frac{b'}{a'} \tag{31}$$

For the outer tube

$$\Phi_r'=2I\log\frac{b'}{r}+\frac{4\pi}{K'^2}\left[-Z_0'(K'x)\right]_r^{b'}-\frac{4\pi a'Z_1'(K'a')}{K'}\log\frac{b'}{r} \tag{32}$$

The voltage drop in the outer tube is

$$e'=\rho' i_r'+\frac{\partial\Phi_r'}{\partial t}.$$

After substituting and simplifying

$$e'=\rho'\left\{\frac{2j\omega I}{\rho'}\log b'+K'a'Z_1'(K'a')\log b'+Z_0'(K'b')-\log r\left[\frac{2j\omega I}{\rho'}+K'a'Z_1'(K'a')\right]\right\} \tag{33}$$

Since e' is not a function of r, the coefficient of log r in eq 33 must vanish, this gives

$$Z_1'(K'a')=\frac{K'}{2\pi a'}I \tag{34}$$

This is one relation fixing the coefficients in Z_0'. The other relation is in eq 28. Substituting eq 34 in eq 28 gives

$$Z_1'(K'b')=0. \tag{37}$$

Voltage drop in inner tube is, similarly

$$e=\rho Z_0(Kb)+Ka\rho Z_1(Ka)\log b-Ka\rho Z_1(Ka)\log r+2Ij\omega\log\frac{b'}{b}+\rho'[Z_0'(K'b')-Z_0'(K'a')]+\rho'K'a'Z_1'(K'a')\log\frac{b'}{a'} \tag{36}$$

Coefficient of terms involving r must be zero, therefore

$$Z_1(Ka)=0 \tag{35}$$

Putting eq 37 in eq 27

$$Z_1(Kb)=\frac{KI}{2\pi b} \tag{38}$$

Inserting eq 37 and 34 in eq 36 gives

$$e=\rho Z_0(Kb)-2Ij\omega\log\frac{b}{a'}+\rho'[Z_0'(K'b')-Z_0'(K'a')] \tag{39}$$

The effective four-terminal impedance for a shunt with the potential lead inside the inner tube is

$$z=\frac{e-\frac{\partial}{\partial t}\Phi_0}{I}$$

From eq 31, 34 and 37

$$\frac{\partial \Phi_a}{\partial t} = \rho Z_0(Kb) - \rho Z_0(Ka) + \rho'[Z_0'(K'b') - Z_0'(K'a')] - 2j\omega l \log \frac{b}{a'} \tag{40}$$

Subtracting eq 40 from 39

$$e - \frac{\partial}{\partial t}\Phi_a = \rho Z_0(Ka) \text{ or } z = \frac{\rho Z_0(Ka)}{l} \tag{41}$$

From eq 24 and 34

$$Z_1'(K'a') = \frac{K'}{2\pi a'},\ I = C_1'J_1'(K'a') + C_2'N_1'(K'a') \tag{42}$$

From eq 24 and 35

$$Z_1'(K'b') = 0 = C_1'J_1'(K'b') + C_2'N_1'(K'b') \tag{43}$$

From eq 23 and 37

$$Z_1(Ka) = 0 = C_1J_1(Ka) + C_2N_1(Ka) \tag{44}$$

From eq 23 and 38

$$Z_1(Kb) = \frac{KI}{2\pi b} = C_1J_1(Kb) + C_2N_1(Kb) \tag{45}$$

Solving eq 42, 43, 44, and 45 for the coefficients

$$C_1' = \frac{K'I}{2\pi a'}\frac{N_1(K'b')}{D'}$$

$$C_2' = \frac{K'I}{2\pi a'}\frac{J_1(K'b')}{D'},$$

where $D' = J_1(K'a')N_1(K'b') - J_1(K'b')N_1(K'a')$,

$$C_1 = -\frac{KI}{2\pi b}\frac{N_1(Ka)}{D} \tag{46}$$

$$C_2 = \frac{KI}{2\pi b}\frac{J_1(Ka)}{D}, \tag{47}$$

where $D = J_1(Ka)N_1(Kb) - J_1(Kb)N_1(ka)$.

From eq 23

$$Z_0(Ka) = C_1J_0(Ka) + C_2N_0(Ka)$$

Using this and putting in values of C_1 and C_2 from eq 46 and 47 eq 41 becomes

$$z = \frac{\rho K}{2\pi b}\left[\frac{N_0(Ka)J_1(Ka) - N_1(Ka)J_0(Ka)}{J_1(Ka)N_1(Kb) - J_1(Kb)N_1(Ka)}\right] \tag{48}$$

Using the general relation

$$N_0(Ka)J_1(Ka) - N_1(Ka)J_0(Ka) = \frac{2}{\pi Ka}$$

eq 48 becomes

$$\therefore z = \frac{\rho}{\pi^2 ab}\frac{1}{J_1(Ka)N_1(Kb) - J_1(Kb)N_1(Ka)} \tag{49}$$

Using the following general relations

$$J_1(Ka)=\sqrt{\frac{2}{\pi Ka}}\left[P_1(Ka)\sin\left(Ka-\frac{\pi}{4}\right)+Q_1(Ka)\cos\left(Ka-\frac{\pi}{4}\right)\right]$$

$$N_1(Kb)=\sqrt{\frac{2}{\pi Kb}}\left[-P_1(Kb)\cos\left(Kb-\frac{\pi}{4}\right)+Q_1(Kb)\sin\left(Kb-\frac{\pi}{4}\right)\right],$$

letting $b-a=d$, and putting in the following approximate values of P_1 and Q_1

$$P_1(Ka)=1+\frac{15}{128K^2a^2}$$

$$Q_1(Ka)=\frac{3}{8Ka}$$

the denominator of eq 49 becomes, for $K^2=-j\mu^2$

$$D_z=\frac{\sqrt{2ab}(1-j)\pi}{m}\left\{\left[\left(1+j\frac{15}{128m^2a^2}\right)\left(1+j\frac{15}{128m^2b^2}\right)+j\frac{9}{64abm^2}\right]\sin\frac{md}{\sqrt{2}}(1-j)+\right.$$

$$\left.\left[\left(1+j\frac{15}{128m^2a^2}\right)\left(\frac{3+3j}{8mb\sqrt{2}}\right)-\left(1+j\frac{15}{128m^2b^2}\right)\left(\frac{3+3j}{8ma\sqrt{2}}\right)\right]\cos\frac{md}{\sqrt{2}}(1-j)\right\} \tag{50}$$

Most of the terms inside the brackets in eq 50 are negligible if the frequency is much above 60 cycles per second. If $f=1{,}000$ c/s, $\rho\leqq 40{,}000$ (cgs units), $(b\text{-}a)\leqq 0.06$ cm, and $b=a=1$ (approximately), this equation simplifies to

$$D_z=\frac{\sqrt{2ab}(1-j)\pi}{m}\sin\frac{md}{\sqrt{2}}(1-j) \tag{51}$$

Letting $\frac{md}{\sqrt{2}}=\delta$ and putting eq 51 back in eq 49

$$z=\frac{\rho\delta}{2\pi d\sqrt{ab}}\left[\frac{(\sin\delta\cosh\delta+\cos\delta\sinh\delta)+j(\cos\delta\sinh\delta-\sin\delta\cosh\delta)}{(\sin\delta\cosh\delta)^2+(\cos\delta\sinh\delta)^2}\right] \tag{52}$$

where

ρ = resistivity of shunt material in cgs units
a = inner radius of shunt tube in cm
b = outer radius of spread tube in cm
d = thickness of shunt-tube wall in cm
$\omega=2\pi$ times the frequency of the current through the shunt.
z = vector impedance of shunt for values of above quantities used.
The d-c resistance of the tube in absohms, for $d<b$ is

$$R_{dc}=\frac{\rho}{2\pi d\sqrt{ab}} \tag{53}$$

and eq 52 becomes

$$z=R_{dc}\delta\left[\frac{(\sin\delta\cosh\delta+\cos\delta\sinh\delta)+j(\cos\delta\sinh\delta-\sin\delta\cosh\delta)}{(\sin\delta\cosh\delta)^2+(\cos\delta\sinh\delta)^2}\right] \tag{54}$$

Equation 54 is very nearly correct for any thin-wall tube when the frequency is above 1,000 cycles per second.

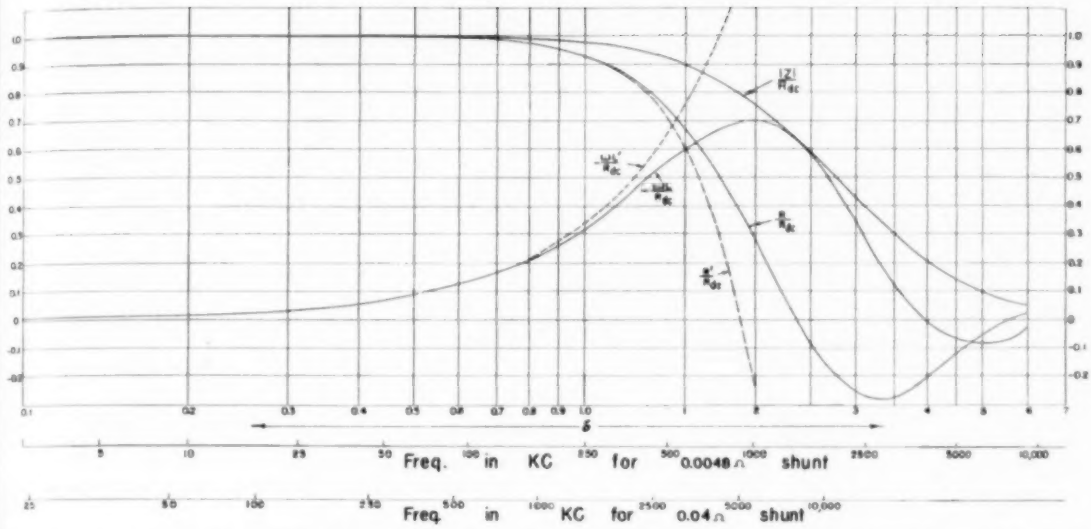


FIGURE 10.—*Curves showing ratios of a-c resistance R, reactance ωL, and impedance Z, to d-c resistance R_{dc}, plotted as ordinates against the parameter, δ, as abscissa.*

Abscissa scales in frequency are also shown for two tubular shunts. The solid-line curves were obtained from the formula derived in appendix 1. The dashed-line curves from formulas known to be applicable only at lower frequencies.

For lower frequencies, the skin effect formula as given on page 91 of Research Paper No. 133 (see reference 5) is

$$R=R_{dc}\left(1-\frac{7m^4d^4}{360}\right)=R_{dc}\left(1-\frac{7}{90}\,\delta^4\right) \qquad (55)$$

The inductance at lower frequencies, where the current is assumed to be uniformly distributed as given on p. 400 of Scientific Paper S281 (see reference 4) is, in abhenries per centimeter,

$$L=-\frac{d}{3a}.$$

For $d<<b$ this gives

$$\frac{\omega L}{R_{dc}}=-\frac{\delta^2}{3} \qquad (56)$$

Actual values of R/R_{dc}, $\omega L/R_{dc}$, and $|z|/R_{dc}$ as functions of δ are shown by the curves in figure 10. The solid lines represent values obtained from eq 54 and the broken lines values from eq 55 and 56. For $\delta<1$, the broken and solid curves coincide, but for $\delta>1$ the values obtained from eq 55 and 56 exhibit rapidly increasing errors as would be expected from the assumptions made in the derivation of these equations.

Frequency scales are also given on these curves for the two shunts in use at this laboratory.

2. Magnetic Force Due to Current in a Tube

Referring to figure 9, the magnetic field (H_r) at any point r distance from the axis of the tube, where $a<r>b$, for a total current, I, in abamperes uniformly distributed over the cross section of the tube, is

$$H_r=\frac{r^2-a^2}{b^2-a^2}\,\frac{2I}{r}.$$

The current density inside the metal of the tube is

$$i=\frac{I}{\pi(b^2-a^2)}.$$

The total current in the shell of thickness dr is

$$i_r=i2\pi r dr=\frac{2rI}{b^2-a^2}\,dr.$$

The total radial force (directed inward) on this shell of current, due to the magnetic field H_r is

$$F_r=i_rH_r$$

in dynes per centimeter length of tube. The total force on the entire tube per centimeter of length is

$$F=\int_a^b F_r=\int_a^b \frac{r^2-a^2}{(b^2-a^2)^2}\,4I^2dr$$

Performing the integration

$$F=\frac{4I^2}{(b^2-a^2)^2}\left(\frac{b^3}{3}-a^2b+\frac{2a^3}{3}\right)$$

This force expressed as a pressure per square centimeter is

$$P=\frac{F}{2\pi b}=\frac{2I^2}{\pi b(b^2-a^2)^2}\left(\frac{b^3}{3}-a^2b+\frac{2a^3}{3}\right),$$

which simplifies to

$$P=\frac{2I^2}{3\pi b}\left[\frac{1}{b+a}+\frac{a}{(b+a)^2}\right],$$

which is the pressure in dynes per square centimeter.

The collapsing pressure for thin-wall tubing as obtained from A. E. H. Dove's formula [7] is

$$P_c = \frac{2E}{1-m^2}\left(\frac{t}{D}\right)^3.$$

E is the modulus of elasticity and m is Poisson's ratio for the tube material; t is the wall thickness of the tube and D is its diameter. If E is expressed in dynes per square centimeter, the collapsing pressure P_c will also be in dynes per square centimeter. Equating this to the magnetically induced pressure due to current, I, as obtained above and solving for I, we obtain the current in abamperes at which the tube will collapse

$$I = 1.085\sqrt{\frac{E}{1-m^2}\,\frac{(b-a)^2(b+a)^2}{b^2(b+2a)}}.$$

3. Effective Inductance of a Type I Concentric-Tube Mutual Inductor as a Function of Frequency

In a type I concentric-tube mutual inductor, shown schematically in figure 11, the current, I, whose rate of

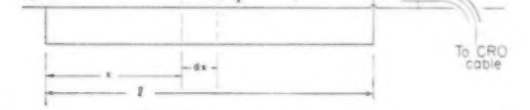


FIGURE 11.—*Schematic drawing of a type I concentric tube mutual inductor.*

change is to be measured, passes through the inner tube and the return circuit for this current is assumed to be far enough away (about 24 in.) so that its effect can be neglected. The secondary circuit of this mutual inductor consists of the length of the inner tube from a to b, the radial ring from b to c, the outer tube from c to d, and the four-wire "spider" from d to a. It is considered to be completed by the resistance, R, which represents the surge impedance of the cable to the CRO. Assuming that (1) all currents flowing in these tubes are uniformly distributed around the central axis, (2) end effects may be neglected, and (3) shunting capacitance from the outer tube to outside conductors may be neglected, differential equations can be written giving the relations of the currents in the tubes to the voltage difference between the tubes at any longitudinal distance, x, from the end of the outer tube. Using symbols defined as follows:

M_1 = mutual inductance per centimeter length of the inner tube, between it and the secondary circuit. Its value is very nearly the same as the self-inductance of the secondary circuit if the wall thickness of each tube is small.

[7] Taken from Marks' Handbook for Mechanical Engineers, Fourth Edition, p. 450.

C_1 = capacitance per centimeter length between the inner and outer tubes,

e_x = the potential difference between the inner and outer tubes at the distance, x, from the end of the outer tube,

i_x = the current in the outer tube at the distance x from the end of the outer tube.

the differential equations are

$$di_x = j\omega C_1 e_x dx \tag{57}$$

and

$$de_x = j\omega M_1(I - i_x)dx. \tag{58}$$

The solution of these differential equations gives equations for i_x and e_x containing two arbitrary constants:

$$e_x = K_1 \sinh\left[\omega x\sqrt{C_1 M_1} + K_2\right] \tag{59}$$

$$i_x = I + jK_1\sqrt{\frac{C_1}{M_1}}\cosh\left[\omega x\sqrt{C_1 M_1} + K_2\right] \tag{60}$$

To obtain values for the arbitrary constants K_1 and K_2 use relations at $x=0$ and $x=l$. At $x=0$, $e_x=0$ and from eq 59

$$0 = K_1 \sinh\left[0 + K_2\right].$$

As K_1 cannot be zero, $K_2=0$.
At $x=l$, $e_x=i_lR$, and from eq 59

$$K_1 = \frac{i_l R}{\sinh \omega l\sqrt{C_1 M_1}}.$$

Putting values of K_1 and K_2 in eq 60

$$i_x = I + j\,\frac{i_l R}{\sinh \omega l\sqrt{C_1 M_1}}\sqrt{\frac{C_1}{M_1}}\cosh \omega x\sqrt{C_1 M_1}$$

at $x=l$

$$i_l\left[1 - jR\sqrt{\frac{C_1}{M_1}}\,\frac{1}{\tanh \omega l\sqrt{C_1 M_1}}\right] = I \tag{61}$$

To get a relation between the voltage impressed on the CRO cable (i_lR) and the rate of change of the current being measured (dI/dt), multiply eq 61 by $j\omega M_1 l$ and take R out of the brackets

$$i_l R\left[\frac{j\omega M_1 l}{R} + \omega M_1 l\sqrt{\frac{C_1}{M_1}}\,\frac{1}{\tanh \omega l\sqrt{C_1 M_1}}\right] = j\omega M_1 l I.$$

As $M_1 l = M$, the total mutual inductance, the above equation becomes

$$\frac{i_l R}{M}\left[\frac{\omega l\sqrt{M_1 C_1}}{\tanh \omega l\sqrt{M_1 C_1}} + j\frac{\omega M}{R}\right] = j\omega I. \tag{62}$$

In this equation i_lR is the voltage impressed on the cable going to the CRO and thus is equal to the voltage recorded by the CRO. The value of M is constant and may be computed from the dimensions of the mutual inductor. The right-hand side of this equation is the rate of change of the current, I, being measured. The quantity in the brackets is unity for all but the higher frequencies. Thus, for most surge work the rate of change of the current being measured is obtained by dividing the voltage recorded at the CRO by the constant M. In order to determine the upper limit of frequency for which this relation holds, the magnitude of the vector quantity in the bracket must be

evaluated as a function of frequency. When the hyperbolic tangent is replaced by the first two terms in its series expansion the magnitude of this quantity becomes:

$$1+\omega^2\left(\frac{M_1C_1l^2}{3}+\frac{M^2}{2R^2}\right).$$

As shown by eq 5 and 6 in the main part of this paper, the product $M_1C_1=1.11\times10^{-21}$ and is the same for any pair of coaxial tubes provided they have thin walls. l is the length of the secondary circuit tube in centimeters. M is the total mutual inductance of the inductor in henries and can be computed from its dimensions as already shown in the main part of this paper. R is the surge impedance in ohms of the cable connected to the potential terminals of the inductor and going to the CRO where it is terminated by a resistance equal to R.

For the type I concentric-tube mutual inductor constructed and used at this laboratory $l=17.8$ cm and $M=0.05\mu$h. The surge impedance of the CRO cable, R, is 50 ohms. For these particular values, the magnitude of the quantity in the brackets of eq 62 varies with frequency as follows:

Frequency	
c/sec.	
30×10^6	1.02
40	1.03_8
50	1.06
60	1.08_5
70	1.11_7
80	1.15_0

For this example of a type I concentric-tube mutual inductor, the rate of change of current as measured at the CRO will be correct to within 10 percent for all frequencies up to about 65 megacycles.

4. Effective Inductance of a Type II Concentric-Tube Mutual Inductor as a Function of Frequency

In a type II concentric-tube mutual inductor as shown schematically in figure 12 the current, I, whose rate of change is to be measured passes through the inner tube (radius r_1) and returns through the outer tube (radius r_3). With this arrangement the current terminals are close together and the self-inductance of the primary current circuit is fixed and quite small. The secondary circuit of this mutual inductor is the same as that of the type I inductor. Making the same assumptions as stated in appendix 3, differential equations can be written giving the relations between currents in the three tubes and voltage differences between the tubes at any longitudinal distance, x, from the end of the middle tube (radius r_2). Using symbols defined as follows:

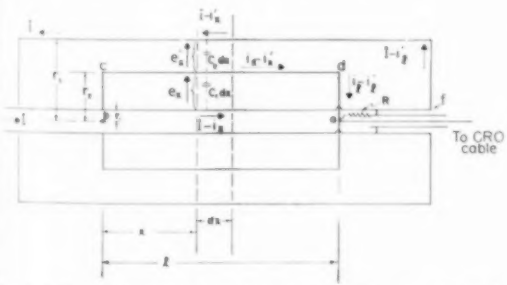


FIGURE 12.—*Schematic drawing of a type* II *concentric tube mutual inductor*

M_1 = mutual inductance per centimeter length of the inner tube, between it and the secondary circuit. This is the same as the self-inductance of the secondary circuit if the wall thickness of each tube is small,

L_2 = self-inductance per cm of the circuit formed by the middle and outer tubes,

C_1 = capacitance per cm between the inner and middle tubes,

C_2 = capacitance per cm between the middle and outer tubes,

e_x = the potential difference between the inner and middle tubes at distance, x, from the end of the middle tube,

e'_x = the potential difference between the middle and outer tubes at distance, x, from the end of the middle tube,

i_x = that part of the current flowing at x in the middle tube whose return path is in the inner tube,

i'_x = that part of the current flowing at x in the middle tube whose return path is in the outer tube,

L_0 = self-inductance of the circuit formed by the inner and outer tubes from a to f. The differential equations are

$$di_x=j\omega C_1e_x\,dx \tag{63}$$

$$di'_x=j\omega C_2e'_x\,dx \tag{64}$$

$$de_x=j\omega M_1(I-i_x)\,dx \tag{65}$$

$$de'_x=-j\omega L_2(I-i'_x)\,dx. \tag{66}$$

The solutions of eq 63 and 65 are

$$e_x=K_1\sinh\left[\omega x\sqrt{C_1M_1}+K_2\right] \tag{67}$$

and

$$i_x=I+jK_1\sqrt{\frac{C_1}{M_1}}\cosh\left[\omega x\sqrt{C_1M_1}+K_2\right] \tag{68}$$

The solutions of eq 64 and 66 are

$$e'_x=K_3\sin\left[\omega x\sqrt{L_2C_2}+K_4\right] \tag{69}$$

$$i'_x=I-jK_3\sqrt{\frac{C_2}{L_2}}\cos\left[\omega x\sqrt{L_2C_2}+K_4\right]. \tag{70}$$

The arbitrary constants K_1, K_2, K_3, and K_4 in the above equations can be determined by using the terminal conditions at $x=0$ and $x=l$.

At $x=0$, $e_x=0$, and $i_x'=0$. So from eq 67

$$0=K_1\sinh\left[0+K_2\right],$$

but K_1 cannot be zero, therefore $K_2=0$. Also, from eq 70

$$0=I-jK_3\sqrt{\frac{C_2}{L_2}}\cos K_4$$

giving

$$K_3=-j\sqrt{\frac{L_2}{C_2}}\frac{I}{\cos K_4}$$

At $x=l$: $e_x=R(i_l-i_l')$ and $e_x'=R(i_l-i_l')+j\omega L_0(I-i_l')$.

So from eq 67, after putting in value of K_2,

$$K_1=\frac{R(i_l-i_l')}{\sinh \omega l\sqrt{C_1M_1}}$$

Also, from eq 69, after putting in value of K_3

$$\tan K_4=-\tan \omega l\sqrt{L_2C_2}+\frac{j\sqrt{\frac{C_2}{L_2}}}{I\cos \omega l\sqrt{C_2L_2}}[R(i_l-i_l')+j\omega L_0(I-i_l')].$$

Using above values of K_1 and K_2 eq 68 at $x=l$ gives

$$i_l=I+jR(i_l-i_l'')\sqrt{\frac{C_1}{M_1}}\frac{1}{\tanh \omega l\sqrt{M_1C_1}} \tag{71}$$

Using above values of K_3 and K_4 eq 70 at $x=l$ gives

$$i_l''=I-\frac{I}{\cos \omega l\sqrt{C_2L_2}}+j\sqrt{\frac{C_2}{L_2}}\tan \omega l\sqrt{L_2C_2}[R(i_l-i_l'')+j\omega L_0(I-i_l'')] \tag{72}$$

Subtracting eq 72 from eq 71 and assuming i_l'' to be very small compared to I in the $j\omega L_0$ term, a value for R $(i-i)$, the voltage across the CRO cable, is obtained

$$(i_l-i_l'')R\left[\frac{\frac{\omega l\sqrt{M_1C_1}\cos \omega l\sqrt{L_2C_2}}{\tanh \omega l\sqrt{M_1C_1}}-\omega M_1 l\sqrt{\frac{C_2}{L_2}}\sin \omega l\sqrt{L_2C_2}+j\frac{\omega M_1 l}{R}\cos \omega l\sqrt{L_2C_2}}{1+\omega L_0\sqrt{\frac{C_2}{L_2}}\sin \omega l\sqrt{L_2C_2}}\right]=j\omega M_1 lI$$

For concentric-tubular construction, if the wall thickness of the tubes is neglected the product of capacitance and inductance per cm is a constant for any two tubes. So let $\sqrt{M_1C_1}=\sqrt{L_2C_2}=a=3.33\times10^{-11}$. Also let $M_1l=M$, the total mutual inductance, then the above equation becomes

$$(i_1-i_l'')R\left[\frac{\frac{\omega al\cos \omega al}{\tanh \omega al}-\omega M\sqrt{\frac{C_2}{L_2}}\sin \omega al+j\frac{\omega M}{R}\cos \omega al}{1+\omega L_0\sqrt{\frac{C_2}{L_2}}\sin \omega al}\right]=j\omega MI$$

By expanding the first term in the numerator of the quantity in the brackets in series form and neglecting higher order terms, the equation becomes

$$\frac{(i_1-i_l'')R}{M}\left[\frac{1-\frac{(\omega al)^2}{6}-\omega M\sqrt{\frac{C_2}{L_2}}\sin \omega al+j\frac{\omega M}{R}\cos \omega al}{1+\omega L_0\sqrt{\frac{C_2}{L_2}}\sin \omega al}\right]=j\omega I \tag{73}$$

This equation is similar to eq 62 of appendix 3 for a type I mutual inductor. Thus, to determine the upper limit of frequency for which a type II inductor can be used, the magnitude of the vector quantity in the brackets of eq 73 must be evaluated as a function of frequency.

A type II mutual inductor having the following dimensions (see fig. 12):

$r_1=0.5$ in.
$r_2=1.0$ in.
$r_3=2$ in.
$l=7$ in.
distance from a to $f=2.5$ in.

is used as an example. By use of eq 5 and 6 in the main part of this paper values for M_1 L_2, C_2, and L_0 may be computed from these dimensions. Then, assuming the CRO cable surge impedance, R, to be 50 ohms, the magnitude of the quantity in the bracket of eq 73 varies with frequency as follows:

Frequency	
c/sec	
10×10^6	0.998
30	.982
50	.952
60	.932
70	.908
80	.885
90	.861
100	.835

For this example of a type II concentric-tube mutual inductor the rate of change of current as measured at the CRO will be correct to within 10 percent for all frequencies up to 70 megacycles.

WASHINGTON, March 28, 1947.

U. S. Department of Commerce
National Bureau of Standards

Research Paper RP1824
Volume 39, September 1947

Part of the Journal of Research of the National Bureau of Standards

Validity of the Cosine-Fourth-Power Law of Illumination

By Irvine C. Gardner

The cosine-fourth-power law states that the irradiance [1] at any point in the image formed by a photographic lens, in the absence of vignetting, is equal to $E_0 \cos^4 \beta$, where E_0 is the irradiance at the center of the field, and β is the angle between the axis of the lens and the conjugate chief ray in the object space. Although not usually so stated, this law involves the additional assumption that the lens is free from distortion. With this assumption, the law applies rigorously when the diaphragm is between the lens and object and the object is at an infinite distance. If the diaphragm is between the lens and the image plane, there may be cases in which the irradiance falls off less rapidly from the center of the field outward than is predicted by the cosine-fourth-power law. A type of negative distortion is defined for which the irradiance is uniform over the entire image. When the diaphragm is within the lens system (the more common condition for photographic lenses), one must know the distortion of the portion of the lens following the diaphragm before a definite statement regarding the irradiance of the image can be made. The departures from exactness of the cosine-fourth-power law arise partly because the effective area of the entrance pupil is a function of the obliquity of the incident chief ray. A method is given for measuring this variation in effective area.

I. Introduction

The cosine-fourth-power law may be stated as follows: In the absence of vignetting, the irradiance (or the effective exposure) for different parts of the image formed by a photographic objective varies as the fourth power of the cosine of the angle between the axis and the chief ray proceeding from the conjugate object area. The desirability of and the economic advantage resulting from the use of photographic objectives of extremely wide angle has made it important to reinvestigate the validity of this law because, unless the consequences of the cosine-fourth-power law can be evaded, negatives made with extremely wide angle photographic objectives will vary to such an extent in density from center to edge that their usefulness is greatly restricted. Although the cosine-fourth-power law is commonly set forth as a precise statement of fact, for most photographic lenses it is, at best, an approximate relation. Slussareff [2] has published a paper of fundamental importance in connection with this subject, in which he demonstrates the importance of the aberrations of the pupils, a characteristic that has been neglected in previous treatments of the cosine-fourth-power law. He concludes that the irradiance of the marginal parts of the field can be increased significantly beyond that predicted by the cosine-fourth-power law by one of the following three optical devices: appropriate aberrations of the entrance pupil; the introduction of negative distortion; and the use of a strongly curved concave image field with a concave photographic plate. Reiss [3] has taken exception to some of the methods of computation employed by Slussareff but accepts the limited validity of the cosine-fourth-power law and the importance of the aberrations of the pupils. In the present discussion the derivations of Slussareff and Reiss will be briefly repeated, and the treatment extended to cases not previously considered.

[1] Throughout the text of this paper the terms *irradiance* and *radiance* as defined by the Optical Society of America (J. Opt. Soc. Am. **34**, 184, 1944) will be used instead of the older terms *illumination* and *brightness*. Illumination and brightness usually relate to radiant energy as evaluated in terms of the luminosity function, a consideration which does not apply when dealing with an image to be recorded by photography. All the relations involved in this paper are geometrical, and hence the equations apply with equal rigor, whether applied to radiant energy, radiation evaluated in terms of the luminosity function, or radiation evaluated in terms of photographic sensitivity.

[2] J. Phys. USSR **4**, 537 (1941).

[3] J. Opt. Soc. Am. **35**, 283 (1945).

II. Magnification Ratios and Distortion of Pupils

If dF is the radiant flux proceeding from one elementary area, dA, to a second, dA' (see fig. 1), its magnitude is given by the equations

$$\left.\begin{aligned} dF &= B\cos\alpha\cos\alpha'\frac{dAdA'}{r^2} \\ &= B\cos\alpha d\Omega' dA \\ &= B\cos\alpha' d\Omega dA' \end{aligned}\right\}, \qquad (1)$$

where B is the radiance of the radiating surface, r is the distance between the two elementary areas, α is the angle between the radius vector and N, the normal to dA, and α' is the similar angle between the radius vector and N', the normal to dA'. In the second and third equations, $d\Omega'$ is the solid angle subtended by the area dA' at dA and $d\Omega$ is the solid angle subtended by dA at dA'. The quantities dA and dA' enter symmetrically in these equations, and dF may be assumed to represent the flux proceeding in either direction between the two elementary areas, provided that the appropriate value of B is used. If the normals to the two elementary areas are parallel, $\alpha=\alpha'$ and eq 1 becomes

$$dF=\frac{B}{e^2}\cos^4\alpha dAdA', \qquad (2)$$

where e is the distance between the parallel planes containing dA and dA'.

In figure 2 a photographic objective is represented with the conjugate areas dA and dA' in the object and image spaces respectively. At S and S', are two additional planes normal to the axis of the lens, one being in the object space and the other in the image space. The shaded pencil includes all rays proceeding from dA through the lens to dA'. The cross sections of this pencil, in the planes S and S' are dS and dS', respectively. If, for the moment, absorption and reflection losses in the lens are neglected, B is the same for each of the four elementary areas under consideration and the flux through all the cross sections of the pencil is the same. Therefore, applying eq 2, one may write

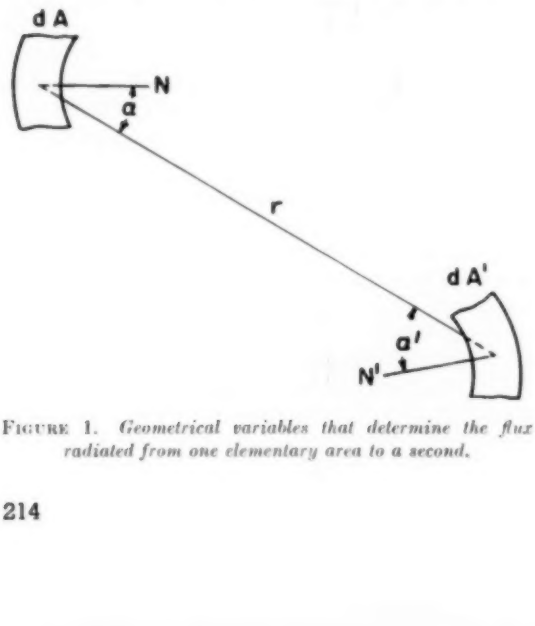


FIGURE 1. *Geometrical variables that determine the flux radiated from one elementary area to a second.*

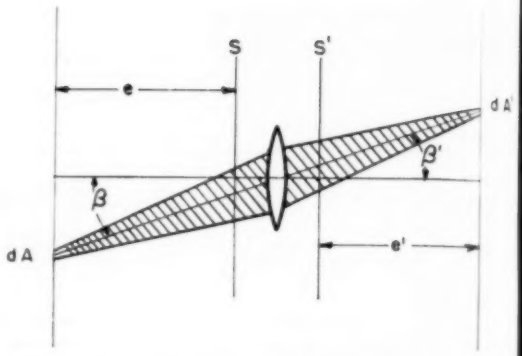


FIGURE 2. *Geometrical variables governing the transfer of radiant energy from the elementary area dA in the object space to the conjugate area dA' in the image space.*

$$\frac{B}{e^2}\cos^4\beta dAdS=\frac{B}{e'^2}\cos^4\beta' dA'dS', \qquad (3)$$

where β is the angle between the elementary pencil in the object space and the normals to the parallel elements of surface dS and dA. In the image space, β' is the corresponding angle. From eq 3

$$\frac{dS'}{dS}=\frac{e'^2}{e^2}\frac{dA}{dA'}\frac{\cos^4\beta}{\cos^4\beta'}. \qquad (4)$$

If it be assumed that the photographic lens is free from distortion, $dA/dA'=1/M^2$, where M is the magnification. Equation 4 can, therefore, be written

$$dS'/dS=C\cos^4\beta/\cos^4\beta', \qquad (5)$$

where $C=e'^2/e^2M^2$ is a constant independent of β and β'. It will now be assumed that S and S' are the planes of the entrance and exit pupils, in which case dS and dS' are conjugate elementary areas in the entrance and exit pupils, respectively.

Equation 5 now has a twofold interpretation. If it be assumed that the conjugate areas dA and dA' remain fixed, the ratio $\cos^4 \beta/\cos^4 \beta'$ will, in general, be different for different pairs of conjugate areas in the entrance and exit pupils. It follows that the exit pupil will be a distorted image of the entrance pupil. If, on the other hand, dS and dS' are assumed to be two conjugate elementary areas that contain the axial points of the planes S and S', the rays making angles β and β' with the axis are chief rays. If now the positions of dA and dA' are allowed to vary within the object and image planes, the ratio $\cos^4 \beta/\cos^4 \beta'$ will, in general, not be constant, and this indicates that the areal magnification at the axial point of the exit pupil, with respect to the entrance pupil, is a function of the inclination of the chief ray. These conclusions are in conflict with deductions based on first-order imagery, and the departures are of the same order as the variations of irradiance in the image plane with which we are concerned. Consequently, it is necessary that these departures from first order imagery not be neglected when computing the irradiance in different parts of the field of a photographic objective.

III. Irradiance at Any Point in Image Plane

If either member of eq. 3 is integrated over the area of the corresponding pupil, one obtains the total flux through the elementary area dA' in the image space. Dividing by dA', one obtains the equation

$$E=\frac{B}{e^2M^2}\int^{S}\cos^4\beta dS=\frac{B}{e'^2}\int^{S'}\cos^4\beta' dS', \qquad (6)$$

where E is the irradiance. For any point in the field of the lens, for which vignetting does not occur, either integration will give the correct value of the irradiance. The angles β and β' are different for different points in the pupil areas S and S', respectively, and are also functions of the point in the field for which the irradiance is being computed. Care must be used in selecting the appropriate boundary for the area over which the integration is extended. If the pupil over which one integrates is an actual physical diaphragm, the integration is extended over the aperture and the limits of integration are the same for all points of the field. On the other hand, if the pupil is an image of a diaphragm positioned somewhere else in the optical system, eq 4 indicates that the area over which the integration is to be performed, in general, will be a function of the point in the field for which the irradiance is sought. The correct area over which the integration should extend is the cross section of the complete pencil lying in the pupil plane. Equation 6 will now be applied to special cases that admit of general treatment.

IV. Special Illustrative Examples

1. Diaphragm Follows the Lens

For this example it is better to integrate the third member of eq 6 because of the simpler limits for the integration. If the aperture of the diaphragm is circular, of radius a, and if β_0' is the angle between the axis and the chief ray passing through the point under consideration, the integral is seen to be identical with that giving the irradiance produced by a disk of radius a and of uniform brightness B at a point distant e' from the disk and at a distance e' tan β_0' from the axis. The exact value of this integral, as given by Foote [4] may be written in the form

$$E=\frac{\pi B}{2}\left[1-\left(1+\frac{4a^2k^2}{(x^2+k^2-a^2)^2}\right)^{-\frac{1}{2}}\right], \qquad (7)$$

where a is the diameter of the disk, k is the distance of the selected point from the plane of the disk, and x is the distance of the point from the axis. These magnitudes and their equivalents in the present notation are shown in figure 3. If θ is defined as tan $^{-1}a/k$ and β_0' is defined as tan $^{-1}x/k$, eq 7 can be written

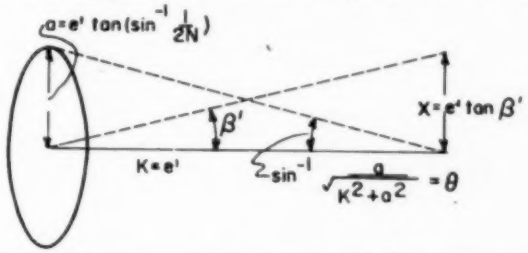


FIGURE 3. *Geometrical quantities involved in computing the irradiance produced by a disk of radius a and of uniform radiance at a distance e' from the disk and at a distance e' tan β' from its axis.*

$$E=\frac{\pi B}{2}\left[1-\left(1+4\cos^4\beta_0'\frac{\tan^2\theta}{(1-\tan^2\theta\cos^2\beta_0')^2}\right)^{-\frac{1}{2}}\right]. \qquad (8)$$

[4] Bul. BS **12**, 583 (1915) S263.

With this notation the aperture ratio of the lens is $1:N$, where $N=1/(2\sin\theta)$. For a point on the axis, $\beta_0'=0$, and eq 8 reduces to the familiar equation

$$E=\pi B\sin^2\theta=\pi B(\tfrac{1}{2}N)^2. \quad (9)$$

If eq 8 is expanded with reference to the exponent $-1/2$ by the binomial theorem and the first two terms of the expansion retained, one obtains the equation

$$E=\pi B\cos^4\beta_0'\left(\frac{\tan\theta}{1-\tan^2\theta\cos^2\beta_0'}\right)^2+\ldots \quad (10)$$

Equation 10 is not recommended for computational purposes, but it does show that the irradiance produced by a uniformly irradiated disk at a point off the axis varies approximately as the fourth power of the cosine of the angular displacement of the point from the axis. By the argument of the preceding paragraph, it follows that when a photographic objective has the diaphragm between it and the image plane, the irradiance of the field varies approximately as the fourth power of the cosine of the angle between the axis and the chief ray in the *image* space.

An estimate of the degree of this approximation is afforded by table 1.

TABLE 1.—*Comparative values of irradiance (or effective exposure) in the field of a photographic objective as a function of relative aperture and angular distance β_0' from the center of field when the diaphragm is between the lens and the image plane*

β_0'	Aperture ratio		
	1:8	1:4	1:2
0°	1.000	1.000	1.000
20°	1.000 $\cos^4 20°$	1.007 $\cos^4 20°$	1.027 $\cos^4 20°$
30°	1.002 $\cos^4 30°$	1.013 $\cos^4 30°$	1.052 $\cos^4 30°$
40°	1.004 $\cos^4 40°$	1.018 $\cos^4 40°$	1.076 $\cos^4 40°$

In this table, the values of the irradiance have been computed by means of eq. 7, which is exact, for aperture ratios 1:8, 1:4, and 1:2 and for the values 0, 20, 30, and 40 degrees. The computed values are expressed as the product of $\cos^4\beta_0'$ and a numerical coefficient. It will be noted that the coefficients of $\cos^4\beta_0'$ are greater than 1 for points not on the axis, indicating that the irradiance of the image falls off less rapidly than $\cos^4\beta_0'$. If a lens is so designed that the second principal plane lies between the lens and the focal plane, the diaphragm may be placed in the second principal plane, and β will equal β_0'. In such a case, one has a lens in which the irradiance of the image decreases from the center of the field outward at a rate somewhat less than that indicated by the cosine-fourth-power law, although the departure is not of sufficient magnitude to be of great importance photographically. It does, however, show conclusively that the cosine-fourth-power law does not represent a limiting condition of maximum attainable uniformity of image irradiance. If the diaphragm precedes the second principal plane, β is greater than β_0', and the departure from the cosine-fourth-power law will be still greater.

2. Diaphragm Precedes the Lens

For this example the first integral of eq 6 is the one more convenient to apply. This integral represents the irradiance produced by a disk of radius b, where b is the radius of the entrance pupil and of brightness B/M^2 at a distance e^2. In most cases encountered in practice, the object will be at a distance of several focal lengths in front of the lens, and consequently the numerical aperture of the incident pencil will be much less (N greater) than the values assumed in the preceding case. Referring to table 1, it is evident that the irradiance will be more nearly proportional to the fourth power of the angle (measured in the object space) than for the preceding example. If, for example, the object is 10 focal lengths' distant, the aperture ratio is approximately $1:10N$. For a 1:2 lens, the aperture ratio in the object space, therefore, becomes 1:20, and it is clear that the departures from the cosine-fourth-power law will be much less than those tabulated for the preceding example.

3. Diaphragm Precedes the Lens, Object at Infinite Distance

This is a particular limiting case of the preceding example. When the first integral of eq 6 is applied, it is discovered that e becomes infinite and M becomes zero in the limit as the distance to the object increases. The limiting value of the product Me is required. If c denotes the distance from the first principal focus to the plane of the entrance pupil,

$$M=f/(e+c) \quad (11)$$

and

$$Me=fe/(e+c). \quad (12)$$

It follows that

$$\lim_{e \doteq \infty} Me = f \tag{13}$$

With the object at infinity, β is constant over the entrance pupil. From eq 6

$$E = \frac{BS}{f^2} \cos^4 \beta, \tag{14}$$

and this example presents an instance in which the cosine-fourth-power law, as stated in the introduction to this treatment, is precisely followed.

4. Diaphragm in Plane of First Principal Focus, Object at Infinite Distance

This is a particular case of example 3. The imagery is telecentric in the image space, and all chief rays are normal to the focal plane. As β_0' equals zero for all parts of the image surface, it has at times been mistakenly assumed that the irradiance will be uniform. It is evident that there is nothing in eq 6 to support this conclusion. The application of the first integral of eq 6 gives the result that has been obtained in the preceding example and indicates the irradiance varies as $\cos^4\beta$. If the second integral is applied, it must be remembered that the exit pupil is infinitely large and at an infinite distance. The value of e' also becomes infinite and the integral remains finite. Although it is true that β_0' is zero for all the pencils, test measurements will indicate that the solid angles included by the pencils at different distances from the center of the field are not equal but vary in accordance with eq 5, and consequently a correct evaluation of the two integrals of eq 6 gives identical results.

5. Diaphragm Precedes Lens, Object at Infinite Distance, Distortion Present

Referring to the first integral of eq 6, when there is distortion the areal magnification, M^2 is a function of β and must remain within the integral. If one writes

$$M^2 = M_0^2 \cos^4 \beta, \tag{15}$$

where M_0^2 is the areal magnification at the center of the field, eq 6 becomes

$$E = BS/e^2 M_0^2, \tag{16}$$

and the field is uniformly irradiated. In the presence of distortion, the linear magnification is not the square root of the areal magnification. For a system possessing rotational symmetry, it can be shown that eq 15 is satisfied, provided

$$r' = M_0 e \sin \beta, \tag{17}$$

where r' is the radial distance from the center of the image plane to the image of an object point distant $e \tan \beta$ from the axis. Applying the condition that e becomes infinite, eq 16 becomes

$$E = BS/f^2, \tag{18}$$

where f is the focal length corresponding to the scale of the image in the neighborhood of the axial point. Equation 17 can therefore be written

$$r' = f \sin \beta. \tag{19}$$

For an undistorted image, the corresponding equation is

$$r'' = f \tan \beta. \tag{20}$$

Consequently, for an object point at the angular distance β from the axis, the linear distortion is $f(\sin \beta - \tan \beta)$, which is negative and corresponds to "barrel shaped" distortion. The ratio of radial magnification of corresponding points in the distorted and undistorted images is $dr'/dr'' = \cos^3\beta$. In general, when negative distortion is present, even though it does not follow eq 16 precisely, the irradiance over the field is more nearly uniform than when distortion is absent.

6. Diaphragm Between Components of Lens

Most photographic objectives are designed with the diaphragm within the optical system, and eq 6 cannot be applied in the general manner of the preceding examples because neither the entrance nor exit pupil is a physical diaphragm, and the region over which either integration must be extended is a function of the position chosen in the image field. It will be assumed that the lens system is divided into two parts, a and b of focal lengths f_a and f_b, the subscript a referring to the part of the lens between the object and diaphragm, and b to the part following the diaphragm. Equation 6 will be applied to the part of the lens consisting of the diaphragm and the part b. The equation becomes

$$E = B/e^2 \int^{S_a} \frac{\cos^4 \beta_d}{M_b} dS, \tag{21}$$

where S_d refers to the area of the diaphragm, β_d is the angle between a ray and the lens axis in the space occupied by the diaphragm, and M_b is the areal magnification for the part of the system under consideration. (It is assumed that the image formed by the first part of the system, which serves as the object for the second part, is stigmatic.) If the rays passing through the diaphragm and proceeding to any given point in the image field are parallel, β_d is a constant over the area of the diaphragm, and eq 21 becomes

$$E=B/e^2 \cos^4 \beta_d \int^{S_d} \frac{1}{M_b} dS. \qquad (22)$$

Even if the rays passing through the diaphragm are not parallel, for most lens systems in use, the intermediate image is remote, and the aperture ratio of the beam is small. Consequently, as is shown by reference to table 1, eq 22, while not rigorously exact, will be a good approximation.

The areal magnification M_b will be a constant if part b of the lens system introduces no distortion. In general, however, part b is a positive lens system preceded by a diaphragm, and the distortion for such a system is frequently negative. As has been mentioned, the effect of negative distortion is to increase the irradiance of the peripheral part of the image as compared with the central part. For most photographic systems β_d is greater than β. Consequently, the factor $\cos^4 \beta_d$ indicates a more rapid decrease of irradiance from the center outward than would be predicted by the factor $\cos^4 \beta$. Therefore, in eq 22, the factors $1/M_b$ and $\cos^4 \beta_d$ indicate departures from the cosine-fourth-power law in opposite directions, and a system with an internal diaphragm must be carefully analyzed before it is possible to say in which direction the irradiance deviates from the cosine-fourth-power law. In particular, if the portion of the lens that follows the diaphragm is free from distortion, M_b is a constant and, as is shown by eq 22, the irradiance varies as the fourth power of the angle between the chief ray and the lens axis in the space occupied by the diaphragm.

V. Effect of Light Losses Introduced by Reflection or Absorption

In the preceding examples it has been assumed that the lens is free from losses by reflection and absorption. In an uncoated lens system such losses are seldom less than 30 percent and may be considerably greater. Actually, the losses by reflection and absorption for any element of a lens system are functions of the angle at which the pencil of light is incident. These losses, therefore, are functions of β or β' of eq 6, and if they are to be correctly taken into account, the transmission coefficient should be within the integral sign. Deriving an analytical expression for the loss by absorption or reflection as a function of the angle of incidence, involving the curvatures of the several surfaces and the variations in thickness of the lens elements introduces so many complications, however, that it is usual to consider the transmittance as independent of β or β' and to write it in front of the integral sign, as a factor multiplied into B. This gives results satisfactorily accurate for most photometric questions that arise in connection with a photographic objective. When the transmission factor is written in front of the integral sign of eq 6, the magnitude of the factor does not affect the indicated variations in the irradiance of the field that have been derived in the several examples of section IV. Neither does the neglect of the effects of absorption and reflection introduce any error in the geometric relations that have been derived between entrance and exit pupils.

One can use a filter that is denser in the center and becomes more transparent from the center outward. If such a filter is placed some distance from a pupil plane, the transmitted beams irradiating the central parts of the field pass through the denser parts of the filter while the more oblique beams that irradiate the peripheral parts of the field pass through the outer less dense parts of the filter. Such a filter acts selectively to make the irradiance of the field more uniform. Unless one has an abundance of light this is an undesirable method for increasing the uniformity of irradiance because it reduces the average irradiance and therefore increases the length of the required exposure. To compute the effect of such a filter, the analytical expression for its transmittance, as a function of β (or β') should be introduced into the integrals of eq 6.

VI. Experimental Method for Measuring Pupils

When a pupil is the image of a diaphragm formed by a portion of the lens system, its size and shape can be most readily determined by a photographic method. If, for example, the size and shape of the

entrance pupil are to be determined, a camera may be placed in front of the lens of which the pupil is to be measured and the diaphragm photographed through the parts of the lens that come between the actual physical diaphragm and the object space. If the camera system is centered on the axis of the lens and directed axially toward the diaphragm, one obtains the pupil corresponding to $\beta=0$. If, now, the camera remains fixed and the lens under test is rotated through the angle β_0 about an axis in the plane of the entrance pupil and at the right angles to the optical axis of the lens, one obtains the pupil corresponding to $\beta=\beta_0$. The images of the entrance pupils are measured on the resulting negatives by means of a planimeter and multiplied by a suitable scale factor to determine the area of the entrance pupil. The diaphragm aperture can be irradiated to be photographed by a ground glass and lamp placed back of the lens. This method of measurement gives the area of the entrance pupil as affected by vignetting if present. To similarly measure the exit pupil, the diaphragm is photographed through the back component of the lens under test.

The method of the preceding paragraph is satisfactory for most purposes but is not rigorously accurate because, for the oblique positions, different parts of the pupil are at different distances from the lens and are reproduced to different scales resulting in a distortion. This can be eliminated by using a telecentric system as shown in figure 4. The lens for which the entrance pupil is to be determined is shown at A. The entrance pupil, which is the image of the diaphragm formed by the first two elements of the system, is represented at B. Actually, for the case illustrated, the entrance pupil would be within the lens, but it is shown here in front of the lens to emphasize its position in the object space. At C there is a lens, preferably corrected for chromatic aberration, which has the small diaphragm D in its second focal plane. Lens G and image plane H schematically indicate the camera with which the photograph is made. The diaphragm at D should in reality be the entrance pupil of the lens G. In the object space of lens C the chief rays are parallel, and therefore the different parts of the entrance pupil are reproduced to the same scale. This arrangement gives a correct determination of the entrance pupil when the object is at an infinite distance.

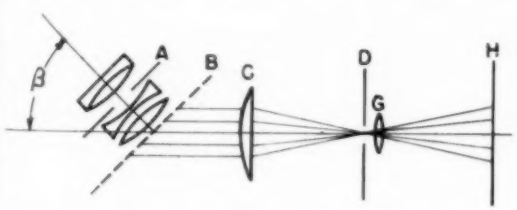


FIGURE 4. *Diagrammatic sketch of arrangements of parts for experimental determination of the dimensions of an entrance pupil of a photographic objective.*

VII. Computational Method for Determining Diameters of Pupils

To determine the size of either entrance or exit pupil by computation is a tedious process. If the entrance pupil is to be determined for the angle β_0, a point in the object plane distant $\epsilon \tan \beta_0$ from the axis is chosen, and rays from this point are traced into the lens as far as the particular space in which the physical diaphragm is located. By successive tests a series of rays are finally found that intersect the edge of the diaphragm. The tracing must be done trigonometrically and, skew rays are required. After a sufficient number of rays intersecting the edge of the diaphragm have been traced, one can construct the surface that encloses all the rays of the transmitted pencil. The cross section formed by the intersection of this surface with the plane of the entrance pupil defines the entrance pupil corresponding to β_0. If there is vignetting, the rays must be traced through the entire system and the size of the maximum transmitted beam determined. To determine the exit pupil, one traces rays in a similar manner, proceeding from the selected image point in the reverse direction through the lens.

WASHINGTON, May 7, 1947.

U. S. Department of Commerce
National Bureau of Standards

Research Paper RP1825
Volume 39, September 1947

Part of the Journal of Research of the National Bureau of Standards

Acid-Base Reactions in Benzene and Other Organic Solvents: Behavior of Bromphthalein Magenta with Different Classes of Organic Bases

By Marion Maclean Davis and Priscilla J. Schuhmann

Practical and theoretical aspects of acid-base reactions in benzene and other organic solvents are discussed. The synthesis of the indicators bromphthalein magenta E and B (the ethyl and *n*-butyl esters of tetrabromophenolphthalein) and of their tetraalkylammonium salts is described. Qualitative and spectrophotometric data concerning the color reactions of the indicators with primary, secondary, and tertiary aliphatic amines, substituted guanidines, and other nitrogen bases in benzene and other types of organic solvents are presented. An explanation offered for the color phenomena is also applicable to sulfonephthaleins and other indicators.

Examples are given of the quantitative measurement of the relative reactivity of bases with bromphthalein magenta in benzene. Illustrations are furnished of the application of the indicators to the quantitative estimation of acids and bases, to measurements of the relative strengths of organic acids, and to the differentiation of primary, secondary, and tertiary aliphatic amines.

During the spectrophotometric investigations, it was necessary to design new types of absorption cells suitable for volatile solvents and for thin layers of solution. A description of these cells is given.

I. Introduction

Practical methods for measuring acidity and basicity in organic media have not kept pace with scientific and industrial needs. Moreover, the fundamental principles governing acid-base reactions in the various types of organic solvents are less understood than the principles governing these reactions in aqueous media and are considerably less familiar to the average chemist. This is understandable, as the great variety of organic solvents now commercially available at moderate cost were for the most part chemical curiosities or completely unknown before World War I. As a result, there has been an overemphasis of certain concepts that were developed in the study of aqueous systems. In particular, the theory that acidic and basic behavior are due, respectively, to hydrogen and hydroxyl ions has been firmly implanted in the minds of most chemists who received their scientific training prior to the last 10 or 20 years. Since then, the more general Brönsted-Lowry concept of acids and bases has gained increasing acceptance. Because an understanding of the work described in this paper requires some familiarity with this theory, a brief explanation will be presented.

The Brönsted-Lowry theory of acidity and basicity.—According to the Brönsted-Lowry theory [1 to 5][1] an acid is defined as a proton donor and a base is defined as a proton acceptor, as indicated in the following relationship:

$$\underset{\text{acid}}{A} \rightleftarrows \underset{\text{base}}{B} + H^+. \qquad (1)$$

Thus, when a compound behaves as an acid it is converted by loss of a proton into a base and, conversely, a base upon combination with a proton becomes an acid. A substance that acts as a base does so because it possesses a pair of unshared electrons. In other words, a base is an electron donor and an acid is an electron acceptor. Later it will be seen that this definition leads to a still more general concept of acids and bases.

Either the acid or the base may be a cation, an

[1] Figures in brackets indicate the literature references at the end of this paper.

755528—47——3

anion, or an electrically neutral molecule, as shown in the following examples:

$$NH_4^+ \rightleftharpoons H^+ + NH_3 \quad (2)$$

$$HC_2H_3O_2 \rightleftharpoons H^+ + C_2H_3O_2^- \quad (3)$$

$$H_2O \rightleftharpoons H^+ + OH^- \quad (4)$$

$$\underset{\text{acids}}{HSO_4^-} \rightleftharpoons H^+ + \underset{\text{bases}}{SO_4^=} \quad (5)$$

Water behaves not only as an acid but also as a base, as shown in the following equation:

$$\underset{\text{base}}{H_2O} + H^+ \rightleftharpoons \underset{\text{acid}}{H_3O^+} \quad (6)$$

It is now generally accepted that the hydrogen ion does not exist in aqueous media as H^+ but is more accurately represented by the formula H_3O^+, as the proton is too readily captured to be capable of independent existence in the presence of a base. This means that reactions such as those shown in eq 2 to 5 will not occur unless a second base is present. For example, a combination of eq 2 and 4 gives

$$\underset{\text{acid}_1}{H_2O} + \underset{\text{base}_2}{NH_3} \rightleftharpoons \underset{\text{acid}_2}{NH_4^+} + \underset{\text{base}_1}{OH^-} \quad (7)$$

of 3 and 6:

$$\underset{\text{acid}_1}{HC_2H_3O_2} + \underset{\text{base}_2}{H_2O} \rightleftharpoons \underset{\text{acid}_2}{H_3O^+} + \underset{\text{base}_1}{C_2H_3O_2^-} \quad (8)$$

and 4 and 6:

$$\underset{\text{acid}_1}{H_2O} + \underset{\text{base}_2}{H_2O} \rightleftharpoons \underset{\text{acid}_2}{H_3O^+} + \underset{\text{base}_1}{OH^-} \quad (9)$$

In all cases, the net reaction is the transfer of a proton from one molecule or ion to another. $Acid_1$ and $base_1$ or $acid_2$ and $base_2$ are termed "conjugate pairs."

It is obvious that any acid dissolved in water will react according to eq 8, just as any derivative of ammonia will react according to eq 7. The uniform behavior of all acids and of ammonia derivatives in dilute aqueous solution is thereby accounted for. Hantzsch, who studied various reactions of acids in different solvents, was the first person to recognize this clearly and he coined the expression "leveling effect" to describe the influence of water in masking differences in the intrinsic strength of acids [6, 7]. Any solvent of basic character, such as liquid ammonia or pyridine, produces a comparable effect, giving NH_4^+, $C_5H_5NH^+$, or in general solvated hydrogen ions. In a similar way, a solvent of acidic character, as for example glacial acetic acid, exerts a leveling effect upon bases [8 to 10].

It becomes evident that solvents can be divided into several categories. They are usually classified in the following way:

1. *Amphiprotic solvents.*—These solvents both give and take up protons with ease. Examples: Water, alcohols. Such solvents can undergo self-ionization, as shown in eq 9. This reaction, termed "autoprotolysis," makes possible the construction of a pH scale for aqueous solutions of acids, bases, and salts.

2. *Aprotic or inert solvents.*—These solvents neither add nor release protons to an appreciable degree. Examples: Benzene and other hydrocarbons; chloroform and other halogenated hydrocarbons. Autoprotolysis and leveling effects do not occur. In these solvents perchloric acid, the strongest acid known, exhibits some tendency to dissociate into its ions, but most acids show practically no ionization. In aprotic solvents, most organic acids exist to some extent as double molecules or "dimers." Bases such as pyridine obviously cannot form ions. It is clear that the term pH, which it will be recalled equals the negative logarithm of the hydrogen-ion activity, is not applicable to solutions of acids or bases in aprotic solvents. However, any pair of acids and bases will react with each other to an extent that depends upon their intrinsic acidic and basic strengths. Reactions of this kind can be followed with the aid of indicator dyes. As is generally known, an indicator dye is either an acid or a base, which in becoming a salt, undergoes a structural modification that is accompanied by a change of color. In some cases, the reaction of an acid and a base can be followed conductimetrically or potentiometrically, as most salts are somewhat ionized even in an inert solvent. These are the methods of greatest practical value, although other chemical methods are known.

3. *Basic solvents.*—Examples are liquid ammo-

nia, aliphatic amines, aniline, pyridine, ethyl ether, dioxane, and esters. Basic solvents exert a pronounced leveling effect upon acids but not upon bases. In some cases, basic solvents exhibit weakly acidic properties and undergo slight autoprotolysis, as for example [11]:

$$2NH_3 \rightleftharpoons NH_4^+ + NH_2^- . \quad (10)$$

4. *Acidic solvents.*—Examples are concentrated sulfuric acid, glacial acetic acid, and other organic acids. These solvents exert a strong leveling effect on bases. Acidic solvents may be weakly basic, and autoprotolysis may occur to a very minor extent, as shown for acetic acid:

$$2CH_3COOH \rightleftharpoons CH_3COOH_2^+ + CH_3COO^- . \quad (11)$$

In fact, Hantzsch attributed the electrolytic dissociation of perchloric acid in an aprotic solvent to a similar reaction, rather than to a dissociation into H^+ and ClO_4^- ions.

The theory of the hydrogen bridge.—A second theory that is vital for the understanding of acid-base reactions is that of the "hydrogen bond," preferably termed the "hydrogen bridge" [12 to 23].[2] This theory postulates that the proton may be shared between two atoms, X and Y, either of the same molecule or of different molecules, to form a bridge between X and Y. In order to be capable of forming a hydrogen bridge, the proton must be part of an acidic group, which means that it is ordinarily attached to a halogen, oxygen, or nitrogen atom. In certain cases, it may be attached to a carbon atom holding a halogen or an unsaturated group, as in chloroform, hydrogen cyanide, or phenylacetylene. Atoms X and Y may be the same or different. For example, there are believed to be OHO bridges connecting the individual molecules of water and the alcohols, an FHF bridge in the anion of hydrogen fluoride, and NHN bridges connecting separate molecules of liquid ammonia. The weakly basic character of an aqueous solution of ammonium hydroxide is attributed to NHO bridges, and the dimerization of organic acids is attributed to OHO bridges. When X and Y are in the same molecule, as in salicylaldehyde, a so-called "chelate ring" is formed. In such cases, the hydrogen bridge may have a high degree of stability.

[2] There is a voluminous literature on the hydrogen bridge. The publications cited were selected mainly because of their general character and because they give references to many other articles on the subject.

Although there can be little question of its widespread occurrence, the nature of the hydrogen bridge is not clearly understood. The proton cannot share two pairs of electrons at one time, but it is possible that a proton originally attached to atom X penetrates to some extent an electron shell of atom Y. The usual explanation is that electronic resonance is set up, giving something intermediate between two extreme forms, such as $-O-H:NH_3$ and $-O^- H:^+NH_3$. The simplest way of representing a hydrogen bridge can be illustrated by ammonium hydroxide: $NH_3-H \ldots OH$. The linkage believed to be the weaker is indicated by a dotted line.

The concept of the hydrogen bridge necessitates a modification of Brönsted's hypothesis. He maintained that, as a consequence of his definition scheme eq 1, a conjugate pair of acids and bases do not possess the same electrical charge, so both cannot be electrically neutral molecules and at least one of them must be an ion. However, as we have just seen, a proton may be only partially surrendered by an acid and only partially captured by the base

The formation of the hydrogen bridge may therefore be regarded as incipient ionization, that is, an intermediate step in the complete transfer of a proton. The bond between hydrogen and atom X or Y is much weaker than an ordinary ionic or covalent bond, and rupture of a hydrogen bridge may lead to formation of either ions or neutral molecules. A familiar example is ammonia water, in which the following equilibria may be postulated:

$$NH_3 + HOH \rightleftharpoons NH_3H \ldots OH \rightleftharpoons NH_4^+ + OH^- \quad (12)$$

The Lewis theory of acidity and basicity.—G. N. Lewis proposed a much broader concept of acids than that of Brönsted and Lowry [24 to 26]. Lewis stated that acidity and basicity have nothing to do with the presence of any specific element or group of elements, and that no particular substance is to be regarded as an acid or a base in an absolute sense. A base is defined as an electron donor and an acid as an electron acceptor. Neutralization is the formation of a coordinate covalent bond between the acid and the base. The adjective "coordinate" is inserted merely to indicate that both electrons are supplied by one of the interacting atoms, as the resulting bond is not different from other covalent bonds. The reaction occurs because a molecule of more stable electronic

configuration results. An excellent presentation of the Lewis electronic theory and a discussion of its manifold applications, as well as a review of other acid-base theories, has been published recently [27].[3] The authors of this textbook draw attention to the analogy between the Lewis concept of acid-base reactions and the modern picture of oxidation-reduction reactions. Acids and oxidizing agents are electron acceptors, while bases and reducing agents are electron donors. However, there is a fundamental difference in the two types of reaction. In oxidation-reduction, electrons are transferred from the reducing agent to the oxidizing agent, whereas when an acid reacts with a base it gains only a share in a pair of electrons.

Lewis gave credit to Brönsted and Lowry for overcoming the reluctance of chemists to think of basicity in terms other than the hydroxyl ion. The list of bases is identical for both theories. However, while the Brönsted-Lowry theory limits acidic properties to proton donors the Lewis theory recognizes any electron acceptor as an acid. In addition to the hydrogen acids, the following compounds are included among *Lewis acids:* BCl_3, $AlBr_3$, $SnCl_4$, SO_3, Ag^+, and Cu^{++}. The reaction of ammonia with boron trichloride is an example of neutralization involving one of these acids:

$$H_3N: + BCl_3 \rightleftharpoons H_3N:BCl_3. \qquad (13)$$

The neutralization product is not necessarily a salt, but may be a covalent compound, as in the example cited. Lewis contended that "to restrict the group of acids to those substances which contain hydrogen interferes as seriously with the systematic understanding of chemistry as would the restriction of the term oxidizing agent to substances containing oxygen" [25]. However, he suggested that the term "hydrogen acids" or, for short, "H-acids" might be used to designate those acids that are proton donors. The great usefulness of Lewis's electronic theory of acids and bases in unifying a vast field of chemical reactions is clearly demonstrated in the references cited. Nonetheless, there are practical advantages as well as theoretical justification for considering the hydrogen acids as a special class. Kolthoff [30] has suggested that the Brönsted-Lowry definition of an acid remain unchanged and that the larger group of acids recognized by Lewis be designated as "Lewis acids" or "proto-acids." To the writers, it is simpler and more logical to consider the Brönsted-Lowry acids as a special subclass, to be designated as hydrogen acids or H-acids, as suggested by Lewis. The acids with which the present investigations are concerned belong to this group, and the bases belong to the class of ammonia derivatives.

Object of the investigation.—Owing to the importance of water as a solvent, the behavior of acids and bases has been investigated very extensively in this medium, but relatively few systematic investigations in organic media have been conducted up to the present time. The result has been a laborious development of empirical and often unsatisfactory procedures, as well as a tendency to extrapolate concepts such as pH to systems in which this term is meaningless. Although academic studies have thrown considerable light upon the subject, there is as yet no systematic compilation of the knowledge so far obtained, and there is a great need for further investigation of acid-base phenomena in various solvents. The purpose of the present work is threefold: To aid in publicizing newer theories that are of considerable value in interpreting and predicting the behavior of acids and bases in organic solvents, as well as the experimental data upon which they are founded; to obtain systematic data in regard to acid-base reactions in organic media, particularly in those solvents that are very dissimilar to water such as the aprotic solvents; and to assist in the development of practical methods for measurement of acidity and basicity in organic solvents.

The study of acid-base reactions in any medium may be divided into two phases:

1. *Determination of the relative strengths of acids and bases.*—In aqueous solutions, these are indicated by the pH values for equimolar concentrations of acids or bases. The pH value for an acid, HX, is a measure of the extent to which the base X^-, present in relatively low concentration, can compete for the proton with the base water which is very weak but present in high concentration. As stated above, the term pH is not applicable to aprotic solvents. However, it is possible to ascertain the relative strengths of a series of acids in an aprotic solvent in terms of the extent of their reactions with some reference base, such

[3] A more detailed review of the historical development of concepts of acids, bases, and salts can be found in a monograph by Walden [28] and in a series of articles published by the Journal of Chemical Education [29].

as a basic indicator dye. Similarly, the relative strengths of organic bases in an aprotic medium can be determined by measurement of their reactivity with an acidic indicator dye. An alternative method is to measure the extent to which an acid (base) competes with an indicator acid (base) for a reference base (acid).

2. *Determination of total acidity or basicity.*—This is accomplished by titration with a strong base or acid. The indicator method for detecting the endpoint is usually the simplest and most convenient, provided indicators of satisfactory properties are available. An indicator should not only undergo color changes in the spectral range to which the eye is sensitive but should be readily soluble and stable towards light and chemical reagents other than acids or bases. Moreover, because colored impurities or suspended particles often render one indicator useless, it is desirable to have a substitute with the same acidic or basic strength but with a different color change. Considerable effort has been expended by a number of investigators in the development of indicators for aqueous media, but only a few of these indicators are suitable for organic media. Before acid-base measurements of organic solutions can be performed as readily as those of aqueous solutions, it will be necessary to provide a series of indicators, ranging from strongly acidic to strongly basic compounds. A further need is for stable, crystalline organic acids and bases sufficiently strong to neutralize weak bases and acids. Work along these lines by the writers is in progress, and it is hoped that others will direct their efforts toward the same objective.

The investigations reported in this paper included the synthesis of the indicator, bromphthalein magenta, and a study of its reactions with different types of organic bases in benzene and other organic solvents. Interesting and unexpected color phenomena were observed, which at first were very puzzling. A consistent explanation has been found that is in harmony with the theories just reviewed, as well as with current explanations for the relation of color to chemical constitution, and is applicable to other indicators. It is believed that these studies point to useful applications of the indicator and that they contribute to an understanding of acid-base phenomena in organic media.

II. Apparatus and Method

The absorption spectra were obtained with a Beckman quartz photoelectric spectrophotometer [31], in which the original cell compartment had been replaced by one designed to hold a pair of 3.8-cm (inside diameter) cylindrical cell holders. A photograph of the parts of the cell holder is shown in figure 1, A. In part of the experiments,

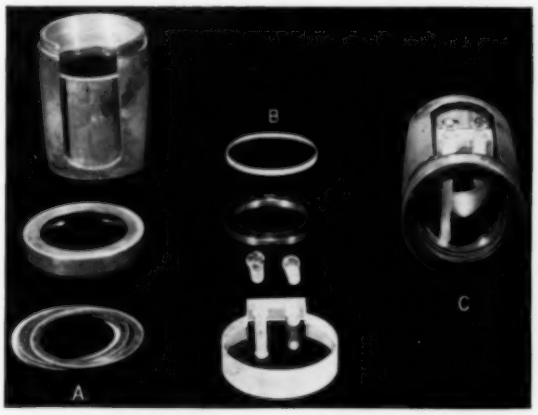


FIGURE 1. *Absorption cell used in spectrophotometric measurements, and its component parts.*

From top to bottom: *A*, Cell holder with metal spacer, threaded cover, gaskets; *B*, quartz endplate, stainless steel spacer, tubulated endplate and stoppers; *C*, cell assembled in the holder.

each cell consisted of a stoppered quartz cylinder 3.8 cm in diameter and 1 cm in length, open and polished at the ends, and fitted with a pair of matched crystalline quartz endplates that were 2 mm in thickness. The cylinder and endplates were assembled tightly in the cell holder by means of a combination of metal spacers and Bakelite and rubber gaskets. To provide a cell 5 mm in length, the quartz cylinder was replaced by a ring of monel metal, gold-plated. This ring (3.8 cm o. d. and 2.8 cm i. d.) was 5 mm in thickness. Two tubular openings through the rim, fitted with glass stoppers, were provided for filling the assembled cell and removing air bubbles.

As a convenient means for obtaining still shorter cells of variable lengths, a new type of tubulated endplate was designed (fig. 1, B, and fig. 2). This was cut in one piece from fused quartz to a thickness of 1 cm, and consisted of a circular plate 3.8 cm in diameter having a neck 2 cm in width and 1 cm in length. Two tubular openings of unequal

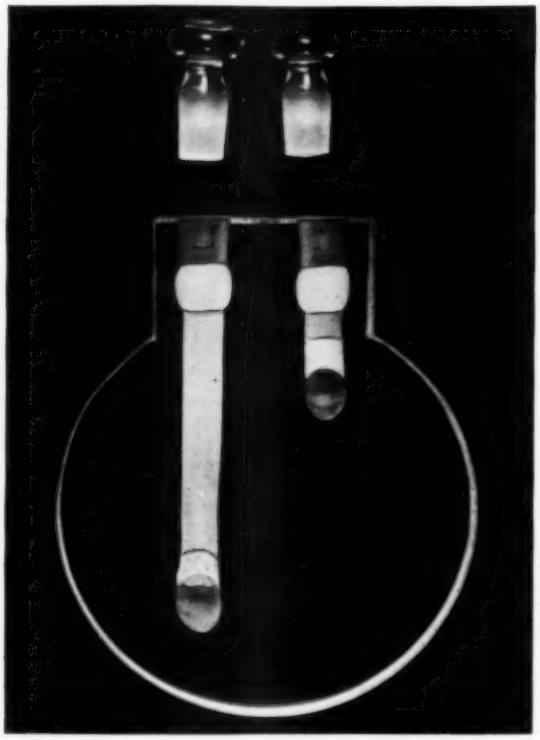


FIGURE 2. *Tubulated quartz endplate with stoppers.*

length were drilled perpendicularly through the neck and continued through the plate so as to emerge at an angle on one face of the plate. The spacing between the openings provided for noninterference with the light beam. Glass stoppers were ground to fit the openings in the neck. When assembled in the cell holder with a stainless steel spacer of the desired thickness and a 2-mm thick crystalline quartz endplate (fig. 1, C), this provided a very satisfactory absorption cell for measurements in the visible spectrum. Two such cells that contained either water or benzene were well matched between 390 and 700 mμ, but did not show equal absorption in the ultraviolet region. When compared with a cell having two crystalline quartz endplates, they showed a very narrow absorption band near 240 mμ, which was probably due to a trace of iron in the fused quartz. The intensity of this band was different in the two tubulated endplates. Stainless steel spacers of thicknesses 1, 2.5, and 5 mm were made, to provide absorption cells of these respective lengths.[4] The solution was introduced into the cell through the longer tube.

In all the spectrophotometric measurements, a pair of cells was used in the customary way. Ordinarily the reference cell contained the pure solvent[5] and its companion cell contained the solution under investigation. The solutions were prepared and the measurements were performed in a room kept at 25° C. By passing a stream of air through the cell compartment, the increase in temperature caused by proximity of the light source was limited to the range 26°±1° C. Solutions of known concentrations were prepared and diluted quantitatively. Volumetric flasks and pipettes calibrated at the National Bureau of Standards were used throughout. The microbalance used for the preparation of most solutions of dyes gave a precision of 0.5 to 1.0 percent for weights of the order used. Readings were made at intervals of 5 or 10 mμ. Measurements of the percent transmittancy were usually reproducible within ±0.2 percent.

In some of the figures to follow, the transmittancies in percent and wavelengths in millimicrons are plotted as ordinates and abscissas, respectively. The resulting curves are called "transmittancy curves." In other figures, values for the molar absorbancy index, a_M, are given as ordinates. These were calculated by means of the equation

$$a_M = \frac{-\log_{10} T_s}{b \times M},$$

in which $T_s = T_{soln}/T_{solv}$ = the transmittancy of the solute, b is the depth of the cell in centimeters, and M is the molar concentration of the solution.[6] In the latter figures, both the wavelength in millimicrons and the wave number (reciprocal of the wavelength) in centimeters $^{-1}$ are given as abscissas, at the top and bottom, respectively, of the figures. These curves are referred to as "molar absorbancy curves."

[4] Since completion of the experimental work described in this paper, combination of a tubulated endplate with a thin quartz plate that has a central indented area 0.1 or 0.5 mm in depth has given satisfactory cells of still shorter length.

[5] Exceptions will be noted in the text.

[6] The symbols and terminology used were recently adopted for use in NBS publications and are discussed in NBS Letter Circular LC857 (May 19, 1947). The term "molar absorbancy index" is used in preference to "molar absorption index" or "molar absorption coefficient."

III. Materials

1. Indicator Dyes

(a) Preparation of the ethyl ester of tetrabromophenolphthalein ("bromphthalein magenta E"[7]).

Phenolphthalin was prepared from phenolphthalein by a procedure based upon Baeyer's brief description [32] and was then converted successively to phenolphthalin ethyl ester, tetrabromophenolphthalin ethyl ester, the potassium salt of tetrabromophenolphthalein ethyl ester, and tetrabromophenolphthalein ethyl ester. The entire series of transformations is shown in figure 3.

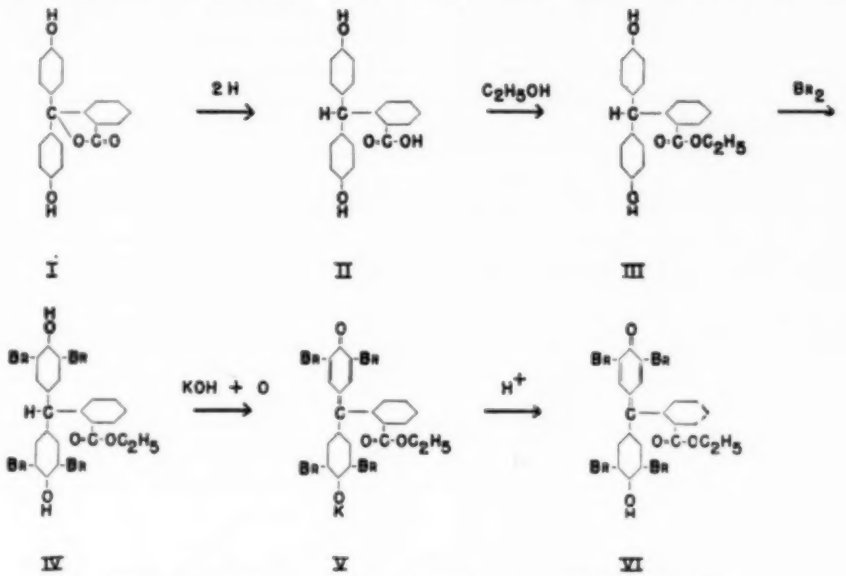


FIGURE 3. *Outline of method for preparing bromphthalein magenta E.*

I, Phenolphthalein; II, phenolphthalin; III, phenolphthalin ethyl ester; IV, tetrabromophenolphthalin ethyl ester; V, potassium salt of tetrabromophenolphthalein ethyl ester; VI, tetrabromophenolphthalein ethyl ester ("bromphthalein magenta E.")

The last four steps were performed essentially by the process outlined by Nietzki and Burckhardt [33]. However, experiments with several small batches were conducted to determine the most favorable conditions, and at various stages modifications were introduced. The following description applies to the procedures found to be most satisfactory.

Preparation of phenolphthalin (II, fig. 3).—Zinc dust (0.76 mole) was "activated" by stirring for 15 minutes with 100 ml of a 25-percent solution of ammonium chloride. The ammonium chloride solution was then decanted, and the zinc was washed twice with distilled water. The activated zinc was added to a 500-ml round-bottom flask that contained a solution of phenolphthalein (0.25 mole, Eastman No. 202) in 125 ml of 10-*M* sodium hydroxide, and the mixture was heated to boiling while constantly stirred. Near the boiling point, the solution became colorless. Heating was continued on the steam bath for an hour. After the addition of 150 ml of 6-*N* sodium hydroxide, the mixture was filtered through asbestos to remove the excess of zinc. The residue of zinc was heated to boiling with 6-*N* sodium hydroxide, and the mixture again filtered through asbestos. The combined filtrates were then poured into a slight excess of 3.6-*N* hydrochloric acid (500 ml). The flocculent precipitate was digested with water on the steam bath and then washed several times by decantation. Finally, it was filtered with the aid of suction and dried in a vacuum oven at 100° C. The yield was 77 percent. The product (mp, 232° to 236° C)[8] may be used without further purification or recrystallized

[7] This name was coined as a substitute for the cumbersome chemical terminology and is intended to suggest the relation of the compound to phenolphthalein and to the sulfonephthalein indicators. The suggested abbreviation is BPM. The suffixes E and B denote the ethyl and *n*-butyl esters, respectively.

[8] All melting points were determined with an ASTM thermometer for 3-in. immersion, calibrated at the National Bureau of Standards.

from *n*-butyl alcohol (about 50 ml of hot butyl alcohol to 10 g of phenolphthalin). The recrystallized product melted at 235° to 235.5° C. Baeyer reported the melting point of phenolphthalin to be 225° C, but in catalog No. 35 of Eastman Organic Chemicals it is given as 232° to 236° C.

Preparation of phenolphthalin ethyl ester (III, fig. 3).—Phenolphthalin (0.2 mole) was dissolved in 1 liter of absolute ethyl alcohol, the solution saturated with dry hydrogen chloride, and the flask closed tightly with a glass stopper. After the solution had stood for several days at room temperature, it was concentrated to 400 ml on the steam bath and then poured onto ice, whereupon a curdy solid precipitated. The yield of air-dried product was 100 percent. The solid was dissolved in 100 ml of ether and extracted with small portions of a 10-percent solution of sodium carbonate until the extract was no longer pink. The ether was evaporated and the residue was dissolved in 1.5 liters of hot benzene. Flat needles (mp, 156° to 157° C) were obtained in 82.5-percent yield. After extraction with carbon tetrachloride the residue melted at 157° to 158° C. Nietzki and Burckhardt reported the melting point as 156° to 158° C.

Preparation of tetrabromophenolphthalin ethyl ester (IV, fig. 3).—A solution of phenolphthalin ethyl ester (0.15 mole) in 225 ml of 95-percent ethyl alcohol was cooled to 0° C, and was then shaken constantly during the dropwise addition of bromine (0.6 mole). The temperature of the mixture was not allowed to rise above 15° C. The mixture was placed in a refrigerator for half a day, during which time a large crop of granular yellowish crystals formed. The precipitate was filtered with the aid of suction, and the yellow color was removed by washing with small portions of alcohol. The yield of air-dried product (mp, 161° to 162° C) was 73 percent. A sample recrystallized from 95-percent alcohol containing Darco charcoal gave irregular prismatic crystals that melted at 162.5° to 163.5° C. Additional dark yellow, sticky solid was obtained by dilution of the alcoholic filtrate and washings with about 2 liters of water. Crystallizing this from alcohol produced a solid tinged with light yellow and brought the total yield to 84 percent. Nietzki and Burckhardt performed the bromination in glacial acetic acid and obtained a product that melted at 163° after recrystallization from acetic acid. No yield was reported. Glacial acetic acid and carbon tetrachloride were found by us to be less satisfactory solvents than ethyl alcohol, from which the product separated more readily and in a purer condition.

Preparation of the potassium salt of tetrabromophenolphthalein ethyl ester (V, fig. 3).—Tetrabromophenolphthalin ethyl ester (0.1 mole) was dissolved in a cold solution of 0.4 mole of potassium hydroxide in 200 ml of water, yielding a colorless solution. The solution was mechanically stirred while potassium ferricyanide (0.2 mole in 200 ml of water) was added rapidly. After about 5 minutes, the dark blue slimy precipitate was filtered with suction, and the moist product was dried on a clay plate.[9] The air-dried solid, in small portions, was extracted with 95-percent ethyl alcohol in a Soxhlet apparatus. A dull green, crystalline solid separated from the alcoholic extracts (total yield, 55 percent). A white residue in the thimbles was identified as potassium ferrocyanide. A second extraction of the solid with 95-percent alcohol yielded lustrous green crystals.

Preparation of tetrabromophenolphthalein ethyl ester, "bromphthalein magenta E" (VI, fig. 3).—The potassium salt of tetrabromophenolphthalein ethyl ester was converted into tetrabromophenolphthalein ethyl ester in portions of 5 g, under slight variations of the following procedure. The potassium salt was suspended in 100 ml of water, and acetic acid was added in small portions until, after thorough shaking, the color of the solid had changed to brick red. Benzene (200 ml) was then added and the mixture was heated on the steam bath under reflux. After half an hour, during which the mixture was occasionally shaken, all but a trace of the brick-red solid had dissolved in the benzene. The orange-red benzene layer was separated, partly dried by filtration through filter paper and, while still hot, was divided between two 250-ml Erlenmeyer flasks and placed in a desiccator that contained calcium chloride and vaseline. The desiccator was connected with the water pump. The solution became slightly turbid but cleared after boiling started. As soon as the liquid became concentrated and cooled by the evaporation of solvent, crystals began to form. These consisted of short, silky, light-yellow needles, which at first aggregated in a sheaf-like arrange-

[9] Subsequent batches were centrifuged at this point with considerable saving of time.

ment then changed into fluffy balls that finally filled the liquid. During filtration these crystals lost a considerable part of their bulk and changed to a bright orange powder. On heating for a few hours in a vacuum oven at 85° to 105° C, the color changed to brick-red. The total yield of oven-dried product from 41.5 g of the potassium salt was 85 percent. The products from five separate preparations gave identical absorption spectra when dissolved in benzene. Recrystallization of 31 g from 1.5 liters of benzene gave 91 percent of oven-dried product that had an unchanged absorption spectrum.

Several experiments were performed to determine whether the benzene of crystallization in the air-dried solid was present in stoichiometric proportions. The results were inconclusive, but may indicate the presence of 1 molecule of benzene of crystallization, which is slowly lost at atmospheric conditions. When heated at atmospheric pressure, the orange powder dries to a porous mass that can be ground to a brick-red powder. The solid melts to a deep-red liquid at about 209° to 210° C. Nietzki and Burckhardt obtained stout blood-red prisms (mp, about 210° to 215° C) from alcohol.

Analysis: [10] Calculated for $C_{22}H_{14}O_4Br_4$: C, 39.9; H, 2.1; Br, 48.3. Found: C, 40.5; H, 2.3; Br, 48.2. For the product of Nietzki and Burckhardt the following analysis was reported: C, 39.4; H, 2.3; Br, 49.4.

(b) Preparation of the *n*-butyl ester of tetrabromophenolphthalein ("bromphthalein magenta B")

Preparation of phenolphthalin n-butyl ester.—Phenolphthalin (0.1 mole) was dissolved in 150 ml of hot *n*-butyl alcohol (commercial high grade). The solution was saturated with dry hydrogen chloride and heated on the steam bath for several days. The excess of butyl alcohol was distilled off under reduced pressure, and the residue in the distilling flask was poured onto ice. The thick sirup that formed was converted into a granular solid after continued stirring. The suction-dried precipitate was dissolved in ether, and the ether solution was extracted with 10-percent sodium carbonate solution until the extract showed only a faint pink color. The ether layer was then washed with water and finally evaporated to dryness on the steam bath. When the residue was dissolved in 100 ml of hot benzene, a yellow solution was formed from which small needles precipitated on cooling. The needles were filtered and washed with cold benzene until no color remained. After 4 hours in a vacuum oven at 80° C, 78 percent of product (mp, 131° to 131.5° C) was obtained.

Analysis: Calculated for $C_{24}H_{24}O_4$: C, 76.6; H, 6.4. Found: C, 76.7; H, 6.7.

Preparation of tetrabromophenolphthalin-n-butyl ester.—The same procedure was followed as for the preparation of the corresponding ethyl ester, except that *n*-butyl alcohol was used in place of ethyl alcohol as the solvent for bromination. Phenolphthalin *n*-butyl ester (0.2 mole) gave a 78-percent yield of the tetrabromo-derivative. The melting point (138° to 138.4° C) was raised to 139.5° to 139.8° C by recrystallization from methyl alcohol.

Preparation of the potassium salt of tetrabromophenolphthalein n-butyl ester.—The procedure corresponded to that used for the ethyl ester. The largest quantity used in an experiment was 0.03 mole of tetrabromophenolphthalin *n*-butyl ester. Acetone was used for extraction of the product in a Soxhlet apparatus. Lustrous green crystals were obtained.

Preparation of tetrabromophenolphthalein n-butyl ester, "bromphthalein magenta B."—The method was similar to that used for the preparation of bromphthalein magenta E. The oven-dried product (yield from 0.0275 mole, 72 percent) was bright orange in color and melted and partly decomposed at about 110° C.

Analysis: Calculated for $C_{24}H_{18}O_4Br_4$: C, 41.8; H, 2.6; Br, 46.3. Found: C, 42.2; H, 3.0; Br, 47.4. The absorption spectrum of a 5×10^{-5}-*M* solution of bromphthalein magenta B in benzene did not differ significantly from that of an equimolar solution of bromphthalein magenta E. The synthesis of bromphthalein magenta B has not hitherto been reported.

2. Organic Bases

The best commercial grade of organic bases was used in all cases. Samples of bases that are available only in research quantities were obtained through the courtesy of several manufacturers. A list of the bases, which includes 14 primary aliphatic amines, 18 secondary aliphatic amines, 12 tertiary aliphatic amines, and 6 guanidine deri-

[10] All the analyses were performed by Kenneth D. Fleischer.

vatives, is given in table 1. The source and the method of purification, if any, are indicated in the footnotes at the end of the table. The preparation of tetraalkylammonium salts is described later in the text. Reference will also be made to trimethylamine oxide, certain alkaloids, and guanidine salts.

TABLE 1.—*Source of organic bases and method of purification*

Primary aliphatic amines	Secondary aliphatic amines	Tertiary aliphatic amines
o-Aminodicyclohexyl [a]	2-(Butylamino)-1-phenylethanol [b]	Diethylbenzylamine [c]
2-Amino-2-methyl-1-propanol [d]	*β*-Butylaminopropionitrile [b]	*N*-Ethylmorpholine [b]
2-Amino-1-butanol [d]	Di-*n*-amylamine [a]	Hyamine B [e]
n-Amylamine [a]	Di-*iso*-amylamine [a]	Hyamine H [e]
iso-Amylamine [a]	Di-*n*-butylamine [f]	Methyldiethanolamine [b]
Benzylamine [f]	Dicyclohexylamine [a]	Pyridine [f]
n-Butylamine [f]	Diethanolamine [a]	Tri-*n*-amylamine [a]
iso-Butylamine [a]	Diethylamine [f]	Tri-*iso*-amylamine [f]
sec-Butylamine [c]	Di-(2-ethylhexyl)amine [b]	Tribenzylamine [a]
Cyclohexylamine [f]	2,3-Dimethylpiperidine [a]	Tri-*n*-butylamine [a]
n-Heptylamine [a]	Di-*n*-propylamine [c]	Triethylamine [f]
n-Octylamine [a]	2-(2-Ethylhexylamino)-1-phenylethanol [b]	Tri-*n*-propylamine [a]
n-Propylamine [c]	*N*-(2-Ethylhexyl)-*N*-*α*-methylbenzylamine [b]	Guanidine derivatives
iso-Propylamine [a]	2-(Methylamino)-1-phenylethanol [b]	1, 2 - Dicyclohexylguanidine [g]
Miscellaneous	*N*-Methylethanolamine [b]	1,2-Diphenylguanidine [h]
Trimethylamine oxide [i]	*α*-Methylpiperidine [a]	1,2-Di-*o*-tolylguanidine [i]
	Octylethanolamine [b]	Dodecylguanidine [b]
	Piperidine [f]	1, 2, 3 - Triphenylguanidine [j]
		1,2-Di-*m*-xylylguanidine [k]

[a] Eastman Kodak Co.; no further treatment.
[b] Carbide & Carbon Chemicals Corp.; no further treatment.
[c] Sharples Chemicals Inc.; no further treatment.
[d] Commercial Solvents Corp.; no further treatment.
[e] Röhm and Haas; no further treatment.
[f] Eastman Kodak Co.; dried over sodium hydroxide and redistilled.
[g] American Cyanamid Co.; isolated from hydrochloride, recrystallized from toluene, and dried in a vacuum oven at 100° C, Mp, 180° to 181° C.
[h] Eastman Kodak Co.; recrystallized twice from toluene and dried in a vacuum oven at 75° C; Mp, 147.5° to 147.8° C.
[i] Eastman Kodak Co.; recrystallized twice from benzene and dried in a vacuum oven at 100° to 110° C; Mp, 175° to 175.5° C.
[j] Synthesized from trimethylamine; mp, 94.5° to 95.5° C (dihydrate).
[k] American Cyanamid Co.; no further treatment.
[l] Eastman Kodak Co.; recrystallized once from cyclohexane and dried in air; mp, 144.5° to 145.5°C.

3. Solvents

Benzene.—A high grade of commercial solvent was dried over Drierite and redistilled [34].

Cyclohexane.—A high-grade solvent suppled for spectroscopic work was used without further treatment.

Dioxane.—Commercial solvent was purified according to the method of Eigenberger [35].

Ethyl acetate.—A high grade of commercial solvent was shaken with dilute alkali, washed, dried, and distilled [34].

Ethyl alcohol.—95-percent alcohol was redistilled [34].

Ethylene dichloride (1, 2-dichloroethane).—A high grade of commercial solvent was used. This showed no acidity when freshly opened. If stored in a colorless glass container in indirect daylight, the solvent acquired an acidic reaction in a few weeks. Deterioration was very slow when the solvent was protected from light.

Other solvents used in qualitative tests were high-grade commercial specimens.

IV. Data and Discussion

1. History of Bromphthalein Magenta

The dye referred to as "bromphthalein magenta E" was first synthesized in 1897 [33] in the course of studies of the structures of phenolphthalein and related compounds. In a similar connection, it has been the subject of spectrographic studies by Thiel and Diehl [36], and by Madame Ramart-Lucas and her collaborators [37, 38]. The dye and its potassium salt have been proposed for the detection of native protein [39].

Bromphenol blue, the sulfonephthalein analog of bromphthalein magenta, has been used as an indicator in acid-base studies in hydrocarbons and their halogenated derivatives, such as benzene [40 to 43] and chlorobenzene [44], as well as in acetonitrile [45]. It has limited solubility in benzene (about 10^{-4} mole per liter) and tends to form precipitates with certain amines. Furthermore, two changes in color, namely from colorless to yellow and from yellow to blue, complicate its use as an indicator [43]. It had been thought likely that the ethyl ester of tetrabromophenolphthalein ("bromphthalein magenta E," "BPM-E") would show a higher solubility in hydrocarbons and related solvents, as well as a single color change from yellow to blue or purple upon reaction with an organic base. The first expectation was fulfilled,[11] and the products of its reaction with secondary aliphatic amines or tetraalkylammonium hydroxides give blue solutions in benzene. However, the interesting discovery was made that tertiary aliphatic amines and symmetrical di- or triaryl-substituted guanidines react with bromphthalein magenta to give magenta solutions in

[11] The solubility in benzene is more than 10^{-3} mole per liter.

benzene and like solvents. Primary aliphatic amines display hybrid behavior, giving red-violet solutions whose absorption spectra show resemblances to those of both secondary and tertiary amines. Furthermore, there are characteristic differences in the blue color and absorption spectra of tetraalkylammonium derivatives of bromphthalein magenta, on the one hand, and solutions of the dye to which secondary amines have been added. These and other color phenomena are discussed in following sections. At this point, however, it may be noted that the differentiating behavior of bromphthalein magenta towards various classes of bases in benzene is not unique but is shown by bromphenol blue and other sulfonephthaleins as well as by dyes of several other classes [43].

Replacement of the ethyl radical in the carbethoxy group by a *n*-butyl radical would not be expected to affect the behavior of bromphthalein magenta as an indicator in a noticeable way but might confer increased solubility in hydrocarbons. Bromphthalein magenta B is indistinguishable from bromphthalein magenta E as an indicator, although it is five or more times as soluble in benzene on a molar basis. In cyclohexane, a 10^{-4}-M solution of bromphthalein magenta B could be prepared, but the solubility of bromphthalein magenta E was too small for use in this solvent.

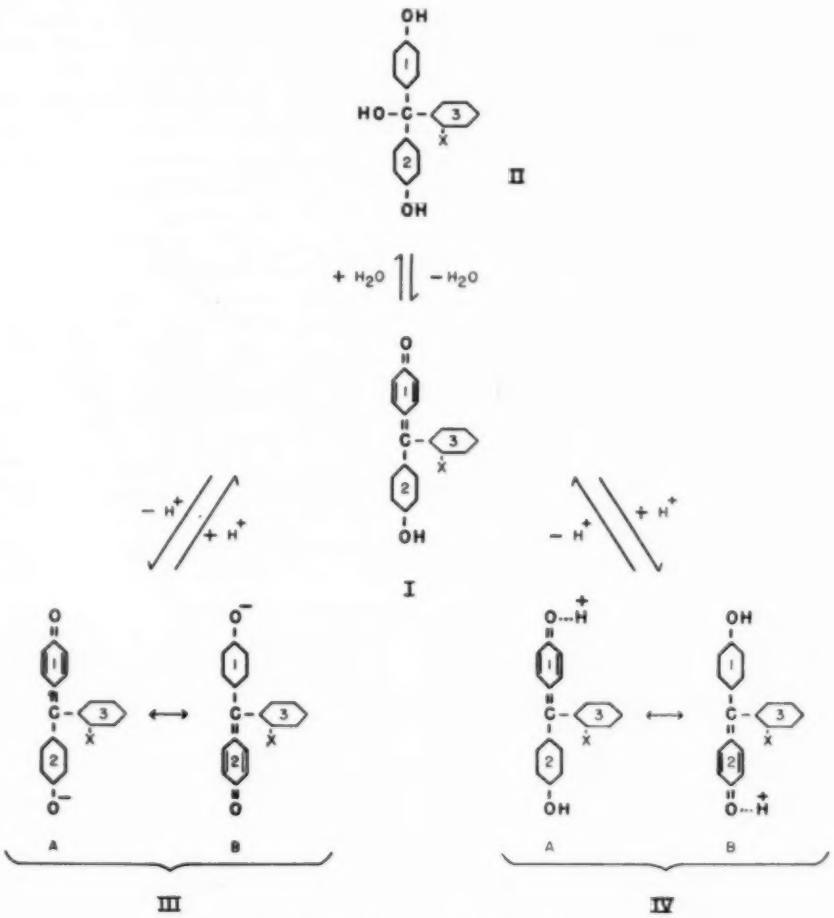


FIGURE 4. *Structural and color changes of benzaurin.*

I, Benzaurin (X=H) (yellow); II, carbinol (colorless); III, anion (deep color); IV, cation (deep color).

2. Structure of Bromphthalein Magenta

Bromphthalein magenta (E or B) belongs to the broad class of triphenylmethane dyes. More specifically, it may be regarded as a derivative of the simplest benzein dye, benzaurin.[12] The formula of benzaurin (I) is given in figure 4. The position marked X in ring 3 is that occupied by $-H$ in benzaurin, by the group $-COOC_2H_5$ in bromphthalein magenta E, and by the group $-COOC_4H_9$ in bromphthalein magenta B. The latter two compounds have 4 bromine atoms in place of the hydrogen atoms ortho to the C—O groups in rings 1 and 2. For simplification of the figure, the bromine atoms have been omitted. The sulfonephthalein dyes are likewise related to benzaurin, having $X=-SO_3H$. Substituents in position X have often been found to protect triphenylmethane dyes from the addition of a molecule of water, which would give the colorless carbinol (II, fig. 4). It is likely, also, that they produce a slight modification of the color of the dyes.[13]

However, the groups that are most actively involved in the structural changes and the accompanying shifts in spectral absorption, which occur when a dye of the benzaurin type is transformed to its anion or cation, are the methane carbon atom and the quinoid (ring 1) and phenolic (ring 2) groups. The simplest explanation of the color phenomena involved is based upon the present-day concept of electronic resonance.[14] The neutral molecule gives yellow solutions in hydrocarbons and halogenated hydrocarbons because the well known chromophoric or "color-bearing" quinoid group $O=C_6H_4=$ is present in the molecule. The anion (III, fig. 4) is formed by the loss of one proton from the $-OH$ group, and the cation (IV, fig. 4) is produced by the addition of one proton to $=O$. In figure 4, it can be seen that III, B, is the same as III, A, except for the distribution of electrons. A similar relationship holds for IV, A, and IV, B. The anion (or cation) is believed to have a hybrid structure, intermediate between the extremes A and B, due to the resonance between these two forms.[15] Such resonance is considered to cause a shift in the absorption of light, from the ultraviolet to the visible portion of the spectrum. When this occurs, the substance in question transmits light of one of the deeper colors, for example red, violet, blue, or green [17, 23, 36, 46 to 50].

The cation form of bromphthalein magenta (red in color) can be obtained by dissolving the dye in concentrated sulfuric acid. The dye in this case is acting as a base. However, its behavior as an acid is of principal interest. It is well known that phenol shows acidic properties that are enhanced by substitution of halogen or nitro groups in the ortho position. This principle was used to advantage in the synthesis of certain sulfonephthalein indicators [51]. An explanation for the acidic behavior of phenol based upon the resonance concept has been given [17, 23]. According to this view, the $-OH$ group lies in the plane of the benzene ring, with the $-H$ directed toward one of the ortho positions. Studies of the infrared spectra of phenol, *o*-bromophenol, 2,4,6-tribromophenol, and other substituted phenols have led to the belief that, when ortho positions in phenol are substituted by halogen atoms, the proton of $-OH$ is displaced to a small extent toward an adjacent halogen atom [52, 53]. When the same ideas are applied to bromphthalein magenta, which has a structural relationship to tribromophenol, its acidic properties may be accounted for and it can be predicted that, because the proton of $-OH$ is attracted toward a neighboring bromine atom within the molecule, the tendency for *inter*molecular association of the $-OH$ groups to give dimers or higher polymers [16] will be absent. In other words, bromphthalein magenta should exist in monomeric form in solution and is therefore likely to conform to Beer's law [54].

3. Color and Absorption Spectra of Bromphthalein Magenta in Different Organic Solvents

Benzene.—Benzene, the solvent for most of the spectrophotometric work described in this paper, was chosen because of its importance as a solvent, availability in high quality at moderate cost, good solvent properties, aprotic character, stability, and miscibility with other organic solvents. In the range of concentrations studied, namely, 1.6×10^{-4} to $3\times10^{-4}M$, the absorption spectra of BPM-E and BPM-B coincided within experimental error and showed no deviations from Beer's law. The

[12] Its relationship to phenolphthalein has already been noted.

[13] This has been found to be true for basic triphenylmethane dyes. Dyes of the benzaurin class have been less thoroughly studied [38].

[14] Equivalent to the British term "mesomerism."

[15] The double-headed arrow is often used to indicate resonance among different structures [23]. A and B are the principal but not the sole contributors to the hybrid anion (or cation).

absorption spectrum of a 10^{-4}-M solution of BPM-E showed no change after 4 months, during which time it was protected only from evaporation and from direct sunlight.

The absorbancy curve obtained for 5×10^{-5}-M bromphthalein magenta E in benzene is shown in figure 5 (curve 1). The high absorption of benzene

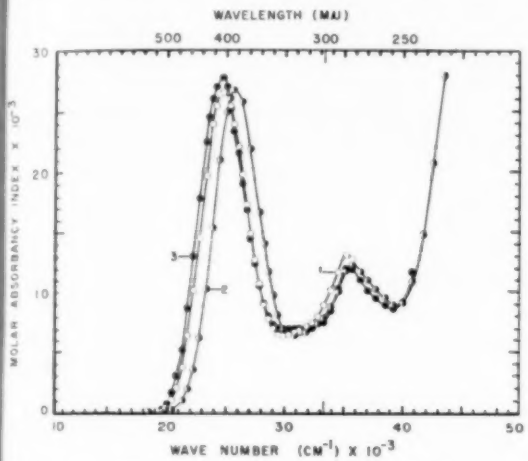


FIGURE 5. *Effect of solvent upon molar absorbancy of bromphthalein magenta.*

(1) Bromphthalein magenta E in benzene; (2) bromphthalein magenta B in cyclohexane; (3) bromphthalein magenta E in ethylene dichloride.

prevents measurements at wavelengths shorter than about 280 mμ, and there is no measurable absorption by the dye at wavelengths greater than about 530 mμ. Two maxima appear near 285 mμ (35,100 cm^{-1}) and 405 mμ (24,700 cm^{-1}), with the respective molar absorbancy indexes 13,000 and 26,500. A minimum is located near 335 mμ (29,800 cm^{-1}), with the molar absorbancy index 6,450.

Ramart-Lucas has given the absorbancy curve for the same compound in benzene [38] with the values for the *logarithm* of the molar absorbancy index plotted as ordinates. When the data for our best sample of bromphthalein magenta E were plotted on the same basis, the absorbancy was found to be greater than that reported by Ramart-Lucas. A sample of bromphthalein magenta E (specimen *A*), obtained from the first trial synthesis, was later shown by its chemical analysis and absorption spectrum to be contaminated with some other substance. The most likely impurity was tetrabromophenolphthalein, which could be formed by saponification of the potassium salt during the synthesis of bromphthalein magenta E. From the chemical analysis, it was calculated that specimen *A* might contain 80 percent of bromphthalein magenta *E* and 20 percent of tetrabromophenolphthalein.[16] This assumption was checked spectrophotometrically; that is, when the absorption spectrum of specimen *A* in benzene was compared with that of a solution that contained the calculated equivalents of bromphthalein magenta E and tetrabromophenolphthalein, the differences in the spectra were within experimental error. The lower values for the absorbancy indexes found by Ramart-Lucas may have been due to the presence of tetrabromophenolphthalein in the ethyl ester used by her. Unlike bromphthalein magenta, tetrabromophenolphthalein does not show visible reaction with organic bases in benzene, and it may therefore be regarded as an inert impurity when bromphthalein magenta is used as an indicator in benzene.

Cyclohexane.—The absorbancy curve obtained for 5×10^{-5}-M bromphthalein magenta B in cyclohexane is given in figure 5 (curve 2). In the visible region, the curve is very similar to that obtained for an equimolar solution in benzene, except that it is slightly narrower and has shifted about 15 mμ toward the ultraviolet. There are two maxima, at the approximate positions 280 mμ (35,700 cm^{-1}) and 390 mμ (25,600 cm^{-1}), with the respective molar absorbancy indexes 12,000 and 27,000. A strong absorption band near 45,000 cm^{-1} is also indicated. Two minima appear near 255 mμ (39,200 cm^{-1}) and 325 mμ (30,800 cm^{-1}), with the respective indexes 9,000 and 6,400. There was no evident deviation from Beer's law in the range of concentrations 1.25×10^{-5} to 5×10^{-5} M.

Ethylene dichloride.—The absorbancy curve obtained for 5×10^{-5}-M bromphthalein magenta E in ethylene dichloride (curve 3, fig. 5) shows absorption and transmission bands in the same regions as those observed for an equimolar solution in benzene but with differences in intensity. The absorption bands found near 285 mμ (35,100 cm^{-1}) and 405 mμ (24,700 cm^{-1}) have the respective molar absorbancy indexes 12,000 and 28,000. That is, absorption of BPM–E in the ultraviolet region is slightly less in ethylene dichloride than in benzene, but in the visible region it is greater. However, these differences in the intensity of

[16] A commercial specimen of the latter compound showed two narrow absorption bands in benzene near 284 mμ and 292 mμ, with the respective molar absorbancy indexes 5,900 and 5,300.

absorption are not great enough to exclude the possibility of their being due to some impurity in the commercial ethylene dichloride. Transmission bands appear near 255 mμ (39,200 cm^{-1}) and 332 mμ (30,100 cm^{-1}), with the respective absorbancy indexes 8,600 and 7,000. The latter transmission band is very flat, showing almost constant transmittancy in the range 310 to 340 mμ.

Other solvents.—Ramart-Lucas obtained nearly identical absorbancy curves for the ethyl ester of tetrabromophenolphthalein in *benzene*, *chloroform*, and *dioxane* [38]. For *alcohol*, she obtained an entirely different curve, with a narrow band of high intensity near 600 mμ. This is because the alcohol reacts as a base with the indicator, just as distilled water acts as a base with bromphenol blue to give the blue form of that indicator.[17] A curve obtained for the potassium salt of bromphthalein magenta E or B in alcohol had a pattern very similar to that obtained by Ramart-Lucas and by the authors for the free indicator acid in alcohol, as well as to a curve obtained by her for a solution of the indicator in aqueous alcohol which contained sodium carbonate. Bromphthalein magenta gives a yellow solution in hydrocarbons and halogenated hydrocarbons other than those mentioned, as well as in very weak bases, such as ethyl ether and dioxane. Acetic acid and other organic acids also give a yellow solution of the indicator. The effect of basic solvents will be discussed more fully in sections IV–8 and IV–10.

[17] Ramart-Lucas does not attribute the change from yellow to blue to reaction with a base, as she does not recognize that alcohol and water have basic properties [38].

4. Reaction of Bromphthalein Magenta with Aliphatic Amines and Symmetrical Di- or Triaryl Derivatives of Guanidine in Benzene

Tertiary aliphatic amines and di- or triarylguanidines.—The yellow color of bromphthalein magenta (E or B) in benzene is changed to a vivid magenta when an excess of a tertiary aliphatic amine or a symmetrical di- or triarylguanidine is added. Solutions in which only part of the indicator has been converted to the "salt" are orange, red, and intermediate tones, depending upon the relative concentrations of indicator and base. The different colors are very bright and easily distinguished by the unaided eye. Examples of organic bases that belong in this group are shown in figure 6. Each of the solutions contained 4×10^{-5}-*M* bromphthalein magenta E, together with 5×10^{-5}-*M* base (1.25 molar equivalents of the indicator) in a 1-cm absorption cell. In numerical order, the transmittancy curves represent the following solutions: (1) Indicator without added base; (2) *N*-ethylmorpholine; (3) 1, 2, 3-triphenylguanidine; (4) tri*iso*amylamine; (5) tri-*n*-butylamine; (6) tri-*n*-amylamine; (7) triethylamine; (8) 1,3-diphenylguandine; (9) 1, 3-di-*o*-tolylguanidine. Other bases not shown in this figure fell into the following respective orders: (1a) Tribenzylamine; (1b) pyridine; (3a) diethylbenzyl-

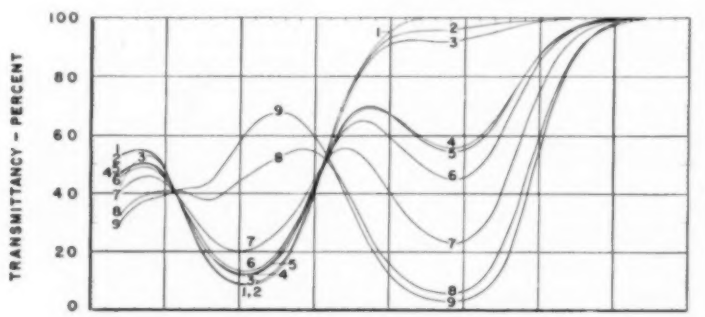


FIGURE 6. *Transmittancy curves for bromphthalein magenta E, 4×10^{-5} M in benzene, in 1-cm cell, with 1.25 molar equivalent of the following organic bases.*

(1) None; (2) *N*-ethylmorpholine; (3) 1, 2, 3, -triphenylguanidine; (4) tri*iso*amylamine; (5) tri-*n*-butylamine; (6) tri-*n*-amylamine; (7) triethylamine; (8) 1, 3-diphenylguanidine; (9) 1, 3-di-*o*-tolylguanidine.

amine; (5a) tri-*n*-propylamine; (10) di-*m*-xylylguanidine. Tribenzylamine and pyridine reacted only feebly. The studies were conducted primarily to determine the general pattern of behavior, and only a single measurement was made with each base. Because of this fact and also because traces of water, carbon dioxide, and in some cases other impurities may have been present in the bases, the quantitive relationships can be regarded as only provisional. However, the extent to which the various bases reacted with bromphthalein magenta gives some indication of their relative strengths. Comparison with the values for relative basic strengths in water is also of interest.

In table 2 are summarized values taken from the literature for the ionization constants in water of some of these bases. The values are expressed as pK_b, that is, the negative logarithm of the ionization constant of the base. When one of these bases is dissolved in water, the following reactions may occur:

$$\underset{\text{base}}{B}+H_2O \underset{\text{(a)}}{\rightleftharpoons} BHOH \underset{\text{(b)}}{\rightleftharpoons} BH^+ + OH^-. \qquad (14)$$

The ionization constant, K_b, refers to reaction (*1b*) and is formulated as follows:[18]

$$K_b=[BH^+][OH^-]/[BHOH]. \qquad (15)$$

Therefore,

$$pK_b=\log\{[BHOH]/[BH^+][OH^-]\}. \qquad (16)$$

In table 2, a smaller value for pK_b indicates a relatively large tendency for ions to form and vice versa. Authorities differ as to the role of reactions (a) and (b) in the evaluation of basic ionization constants. For example, the ionization constant is sometimes defined as $[BH^+][OH^-]/[B]$, or as $[BH^+][OH^-]/\{[BHOH]+[B]\}$. Briegleb [55] has asserted that for the reaction of ammonia with water, reactions (a) and (b) are practically complete, and the ionization constant for ammonia is actually a measure of the equilibrium between $[NH_{3\,aq}]$ and $[NH_4OH_{aq}]$. Moore and Winmill, on the other hand [12, 56], assumed the reaction of B and H_2O to give BHOH, followed by ionization of BHOH, with the respective equilibrium constants K_2 and K_1, and attempted the measurement of K_2 in order to arrive at a true value for K_1. In addition to the resultant uncertainty as to how closely the "apparent" ionization constant in water approximates the "true" ionization constant, the low solubility of some of the organic bases in water is an obstacle to accurate measurements in this solvent [56]. Further considerations have been summarized by Hall [10].

[18] Concentrations are expressed as moles per liter; for simplification, concentrations are assumed to be identical with activities.

TABLE 2.—*Dissociation constants of some tertiary amines and arylguanidines in water* [a]

Compound	pK_b	Relative reactivity with bromphthalein magenta [b]
Pyridine	[c] 8.79; [d] 8.97	1
1,2,3-Triphenylguanidine	[e]. [e] 4.9	2
Diethylbenzylamine	[f] 4.52	3
Tri-*n*-butylamine	[g] 3.17; [e] 4.15 [e] 4.09	4
Tri-*n*-propylamine	[h] 3.26; [f]. [i] 3.35	5
Triethylamine	[f] 3.28; [e] 3.26; [d] 3.18 [h] 3.19; [i] 2.95	6
1,3-Diphenylguanidine	[e] 4.0; [e] 3.94; [j] 4.05 to 4.12	7
1,3-Di-*o*-tolylguanidine	[h] 4.33	8

[a] In some cases, measurements were made in aqueous alcohol, and the value for aqueous solution was obtained by extrapolation. Values are for 25° C, unless otherwise indicated.

[b] In benzene, in increasing order.

[c] N. F. Hall, J. Am. Chem. Soc. **52**, 5115 (1930).

[d] H. T. S. Britton and W. G. Williams, J. Chem. Soc. (London) 796 (1935); values are for 18° C.

[e] N. F. Hall and M. R. Sprinkle, J. Am. Chem. Soc. **54**, 3469 (1932).

[f] J. Kendall, Int. Crit. Tables VI, 260 (National Research Council, Washington, D. C., 1929).

[g] Damsgaard-Sørensen and A. Unmack, Z. phys. Chem. [A] **172**, 389 (1935), value for 18° C; Hall's value 4.15 (see footnote [e]) is believed by these authors to be in error.

[h] G. Bredig, Z. phys. Chem. **13**, 289 (1894).

[i] T. S. Moore and T. F. Winmill, J. Chem. Soc. (London) **101**, 1635 (1912).

[j] P. Walden and H. Ulich, Z. Elektrochem. **34**, 25 (1928).

[k] L. Metz, Z. Elektrochem. **34**, 292 (1928).

When an organic base, B, reacts with the indicator acid bromphthalein magenta, HOR, in benzene, the equation $B+HOR \rightleftharpoons BHOR$ can be assumed for the initial reaction. There is strong evidence that in a partially substituted ammonium salt, "non-Coulombic interaction" occurs between the negative ion, and the proton of the positive ion in nonaqueous solvents [28, 57]. In other words, a hydrogen bridge is formed between the cation and the anion. It seems to be fairly well established that in salts of the type BHOR the hydrogen bridge is strong, even in a solvent of relatively high dielectric constant. The cation and anion are also attracted to each other by a Coulombic force [57]. It therefore seems likely that in benzene the extent of dissociation of BHOR into BH^+ and OR^- ions is negligible in many cases, though the molecule probably has a highly polar character. In benzene, consequently,

the strength of the base is not to be measured in terms of the free ions formed, but by the extent to which B and HOR combine to form the salt BHOR.[19] The basicity constant, K_b, may then be formulated in the following way:

$$K_b=[^+BH^-OR]/[B][HOR]. \quad (17)$$

When the values for pK_b in water of the bases listed in table 2 are compared with the relative reactivities of the same bases with bromphthalein magenta in benzene, there seems to be a direct parallelism for the last five bases. The lack of conformity of the first three bases to this relationship may be due to the fact that their ionization constants in water are not a true measure of the extent of combination of the bases with water. The validity of ionization constants as an index of intrinsic basicity has been questioned by many writers, and the need for further experimental data, particularly for nonaqueous solvents, has been stressed. Furthermore, recent experiments have shown that in some instances steric factors interfere with the estimation of relative basicities [59, 60]. For example, the basicities of tertiary aliphatic amines are lower than would be predicted on theoretical grounds [60].

There evidently is considerable uncertainty in the measurement of basicities in water. The sensitivity of the reaction of organic bases with bromphthalein magenta in benzene demonstrates the possible usefulness of this and other indicators for studying the effect of structure upon intrinsic basic strengths. In fact, the greater solubility of these bases in benzene and the aprotic character of this solvent make the interpretation of data less equivocal than for parallel studies in water.

In figure 7 transmittancy curves are given to show the progressive reaction of 4×10^{-5}-M bromphthalein magenta E in benzene in a 1-cm absorption cell with various increments of 1,3-diphenylguanidine. The smallest increment of base was 0.025 molar equivalent of the dye (10^{-6}-M base), and with less than 3.75 molar equivalents (1.5×10^{-4}-M base) the "limiting curve" was obtained.[20] Two well-marked isosbestic points[21] occur at approximately 355 and 460 mμ. The magenta form gives a broad absorption band with maximum near 540 mμ and a transmission band near 420 mμ. Quantitative aspects of this reaction will be discussed more fully in a later section.

In figure 8 is shown the molar absorbancy curve of the salt that produces a magenta solution in benzene. The compositions of the solutions that gave the three curves shown were as follows: (1) bromphthalein magenta B, 5×10^{-5} M, with 5×10^{-3}-M triethylamine (molar ratio, 1:100); (2) bromphthalein magenta E, 5×10^{-5} M, with 5×10^{-4}-M 1, 3-di-o-tolylguanidine (molar ratio, 1:10); (3) bromphthalein magenta B, 2.5×10^{-5} M, with 2.5×10^{-4}-M ditolylguanidine (molar ratio, 1:10). The concentration of base was in each case

[19] According to some authorities, this should be called a "pseudo-salt."

[20] So termed because continued addition of base produced no further change in the transmittancy curve, showing that reaction of the indicator with the base had reached completion.

[21] Isosbestic means "of equal extinction" [61]. Well-marked isosbestic points are interpreted as showing that two and only two colored species are present in the solution under examination. Isosbestic points at the same spectral regions may be noted in figure 6.

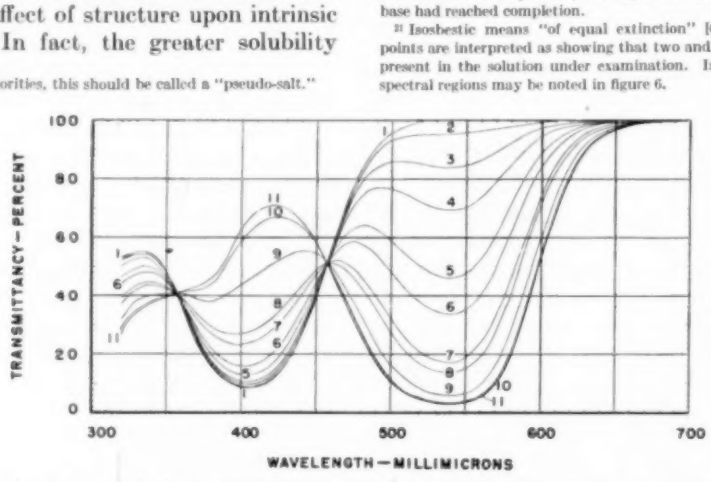


FIGURE 7. *Transmittancy curves for bromphthalein magenta E, 4×10^{-5} M in benzene, in 1-cm cell, with various concentrations of 1,3-diphenylguanidine expressed as molar equivalents of the indicator.*

(1) None; (2) 0.025; (3) 0.0625; (4) 0.125; (5) 0.25; (6) 0.375; (7) 0.625; (8) 0.75; (9) 1.25; (10) 2.5; (11) 3.75 (limiting curve).

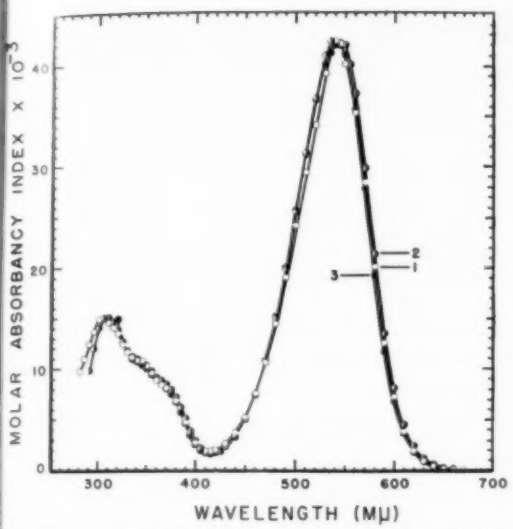


FIGURE 8. *Typical molar absorbancy curves for products from complete reaction of bromphthalein magenta with tertiary aliphatic amines or 1,3-diarylguanidines in benzene.*

(1) Bromphthalein magenta B, 5×10^{-5} *M*, plus 100 molar equivalents of triethylamine; (2) bromphthalein magenta E, 5×10^{-5} *M*, plus 10 molar equivalents of 1,3-di-o-tolylguanidine; (3) bromphthalein magenta B, 2.5×10^{-5} *M*, plus 10 molar equivalents of 1,3-di-o-tolylguanidine.

greater than that required to produce the limiting curve. There are two absorption bands in the spectrum, one near 305 mμ (a_M=approximately 15,000) and a second near 540 mμ (a_M=about 42,000), together with evidence of overlapping bands in the region 340 to 370 mμ and a transmission band near 415 mμ (a_M=2,000). When these curves were obtained, the absorption cells available were not short enough to permit measurements of high accuracy near 540 mμ. This is probably the cause of the slight differences among the curves in this region. In order to conserve base, and also because the effect of the indicator upon the spectral absorption of the base is not known, benzene was used in the comparison cell instead of a solution containing the base. This would account for the small differences in the region 280 to 330 mμ. It is evident that the change in absorption is small when (1) bromphthalein magenta B is used instead of bromphthalein magenta E; (2) ditolylguanidine is used instead of triethylamine; or (3) the concentration of indicator is altered to a limited degree.

It should be mentioned—and this statement applies to all of the reactions of bromphthalein magenta with bases—that in all of the experiments performed, the color change produced by adding one of these bases to this indicator in benzene appeared to be instantaneous. The reaction is exothermic, as shown by the fact that heating the solution causes fading toward the original yellow, and cooling produces a deepening toward magenta. The heat of reaction appears to be smaller for the reaction of bromphthalein magenta with a tertiary amine than for its reaction with primary or secondary amines.

Secondary aliphatic amines.—The yellow color of bromphthalein magenta in benzene is changed to purple-blue ("cornflower" blue) when an excess of a secondary aliphatic amine is added.[22] Solutions that contain free indicator acid mixed with some of its salt with a secondary amine are green or another intermediate color, depending upon the relative amounts of the yellow and blue substances. This is in striking contrast to the behavior of bromphthalein magenta with tertiary amines, which was described in the preceding section. The appearance of the transmittancy curves is illustrated in figure 9. The concentrations of indicator and of base were the same as those that gave the curves in figure 6; that is, each solution contained 4×10^{-5}-*M* bromphthalein magenta E in benzene together with 5×10^{-5}-*M* amine (molar ratio, 1:1.25) in a 1-cm absorption cell. Ignoring for the moment curves 2 to 13, the curves represent solutions containing the following bases: (1) None; (14) di-*n*-propylamine; (15) di-*n*-amylamine; (16) di-*n*-butylamine; (17) diethylamine; (18) di*iso*amylamine; (19) dicyclohexylamine; (20) piperidine. Other secondary amines not given in the figure, in their relative order of apparent strength, are (17a) di-*n*-octylamine; (19a) 2,3-dimethylpiperidine; (19b) 2-methylpiperidine. Dibenzylamine showed negligible reaction, and di-2-ethylhexylamine was much less reactive than dipropylamine. These relative orders are decidedly provisional. In the first place, it is evident from figure 9 that the transmittancies near 575 mμ, the region of maximum absorption, are too close to zero for precise measurements. At the time these measurements were made the shortest available absorption cell was 1 cm long, and it was desired to use concentrations of indicator and

[22] This statement should be qualified as referring to daylight illumination and to relatively dilute solutions or to thin layers of more concentrated solutions. By incandescent light or in thick layers, the color is dark red. This behavior resembles the so-called "dichromatism" of some of the sulfonephthalein indicators, including bromphenol blue [62, 63].

135528—47——4

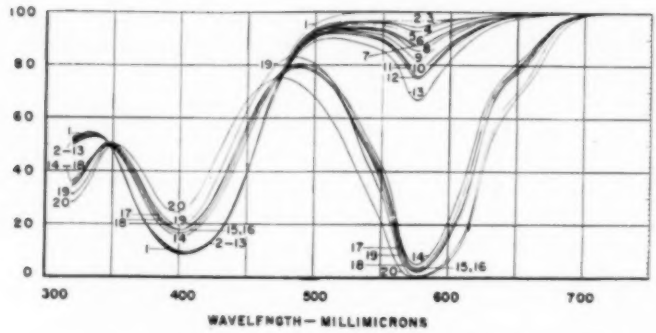


FIGURE 9. *Transmittancy curves for bromphthalein magenta E, 4×10^{-5} M in benzene, in 1-cm cell, with 1.25 molar equivalent of the following organic bases.*

(1) None; (2) benzylamine; (3) *iso*butylamine; (4) *o*-aminodicyclohexyl; (5) *n*-propylamine; (6) *iso*amylamine; (7) *n*-butylamine; (8) *sec*-butylamine; (9) *n*-amylamine; (10) *iso*propylamine; (11) cyclohexylamine; (12) *n*-octylamine; (13) *n*-heptylamine; (14) di-*n*-propylamine; (15) di-*n*-amylamine; (16) di-*n*-butylamine; (17) diethylamine; (18) di*iso*amylamine; (19) dicyclohexylamine; (20) piperidine.

amine that would permit comparisons with other experiments. Variations of a few tenths of a degree in the temperature during measurement are also more serious than in the case of measurements involving tertiary amines. Although the temperature of the room was kept constant without much difficulty, heating of the solutions by the adjacent light source presented a problem. For these reasons and also because several of the amines were available only in small quantities, purification was dispensed with in some cases. However, for the solution that contained di-*n*-butylamine, the same transmittancy curve was obtained irrespective of whether the commercial high-grade amine was used without treatment or after drying and redistilling. It was at first supposed that a trace of water in the secondary amine might be the explanation for a blue rather than a magenta color, but no such effect of water could be demonstrated.

The curves have common intersections near 345 to 350 mμ and 470 to 475 mμ, with the exception of the curve for the solution that contained dicyclohexylamine. The latter curve intersects the curve for the indicator at approximately 480 mμ. The deviation may be characteristic of this amine or may be evidence of an impurity in the sample.

The values for pK_b in water for some secondary aliphatic amines are given in table 3. When these values are compared with those given in table 2 for corresponding tertiary aliphatic amines, the following facts are illustrated: 1, The secondary aliphatic amines, with lower values for pK_b, show greater basicity in water; 2, variations in basicity among secondary aliphatic amines are relatively slight; 3, quantitative comparisons of the basicity of secondary amines are difficult, because of the variation in the values of pK_b for a given amine reported by different investigators.

TABLE 3.—*Dissociation of some secondary aliphatic amines in water* [a]

Compound	pK_b
Diethylamine	[b] 3.00; [c] 3.01; [d] 3.05; [e] 2.94; [f] 2.90; [g] 2.84
Di-*n*-propylamine	[b] 3.09; [c] 3.02; [f] 2.99
Di-*n*-butylamine	[c] 2.79; [d] 2.82
Di-*n*-amylamine	[c] 3.11
Di-*iso*amylamine	[b] 3.11; [c] 3.07; [f] 3.02
Di-*n*-octylamine	[b] 2.99
Piperidine	[b] 2.90; [c] 2.87; [d] 2.91; [f] 2.80; [i] 2.92
2-Methylpiperidine	[c] 3.02

[a] In some cases, measurements were made in aqueous alcohol, and the value for aqueous solution was obtained by extrapolation. Values are for 25° C unless otherwise indicated.

[b] J. Kendall, Int. Crit. Tables **VI**, 260 (National Research Council, Washington, D. C., 1929).

[c] N. F. Hall and M. R. Sprinkle, J. Am. Chem. Soc. **54**, 3469 (1932).

[d] N. F. Hall, J. Am. Chem. Soc. **52**, 5115 (1930).

[e] T. S. Moore and T. F. Winmill, J. Chem. Soc. (London) **101**, 1635 (1912).

[f] G. Bredig, Z. phys. Chem. **13**, 289 (1894).

[g] H. T. S. Britton and W. G. Williams, J. Chem. Soc. (London), 796 (1935); values are for 18° C.

[h] C. W. Hoerr, M. R. McCorkle, and A. W. Ralston, J. Am. Chem. Soc. **65**, 328 (1943); probable accuracy ±0.01.

[i] T. S. Moore, J. Chem. Soc. London **91**, 1373 (1907).

The transmittancy curves in figure 10 show the reaction of bromphthalein magenta E in benzene with progressively increasing amounts of piperidine. The curves correspond to 5×10^{-5} M indicator in a 5-mm cell with the following concentrations of piperidine, expressed as molar equivalents of dye: (1) None; (2) 0.1; (3) 0.3; (4)

0.5; (5) 1.0; (6) 10 (limiting curve). In other words, the reaction was complete upon or before the addition of 5×10^{-4}-*M* piperidine. Three isosbestic points at 295, 345, and 475 mμ are evident.[23] The sensitivity of the reaction is noteworthy. Quantitative aspects will be referred to in a later section.

The molar absorbancy curve of the "salt" formed by the complete reaction of bromphthalein magenta with a secondary aliphatic amine in benzene is shown in figure 11. Curve 1 corresponds to 2.5×10^{-5}-*M* BPM-E with 40 molar equivalents of diethylamine, and curve 2 corresponds to 5×10^{-5}-*M* BPM-E with 10 molar equivalents of piperidine. In both cases an excess of base was present. These two curves are virtually superimposed and except in the nearly vertical portions, where measurements are less precise, values for the two solutions are almost identical. The absorbancy curve shows a very strong band near 575 mμ (a_M=approximately 90,000) and two weaker bands near 315 mμ (a_M=18,000) and 395 mμ (a_M=about 10,000), with transmission bands in the regions 360 mμ (a_M=6,000) and 450 mμ (a_M=less than 1,000). When this absorbancy curve is compared with that for the salt that is formed by reaction of bromphthalein magenta with a tertiary amine (see fig. 8), the striking difference between the two absorbancy curves, that correspond respectively to purple-blue and magenta solutions in benzene, is readily perceived. The difference would be even more apparent had the two graphs been plotted with the same relation between ordinates and abscissas. The maximum molar absorbancy for the purple-blue solution is more than twice that for the magenta solution. The main band obtained for the former solution is very steep and narrow, while that for the magenta solution is relatively much broader. Differences in the spectral region, 280 to 400 mμ, are also obvious. The differences in the reactions of secondary and tertiary amines with bromphthalein magenta in benzene will be discussed more fully in sections IV–8 and IV–10.

Primary aliphatic amines.—The color produced by adding an excess of a primary aliphatic amine to bromphthalein magenta in benzene is a red-purple, sometimes referred to as "plum color."

[23] Piperidine was not present in the comparison cell, but even at the highest concentration used in these experiments its effect upon the transmittancy of benzene is only about 0.3 percent at 280 mμ and less at longer wavelengths.

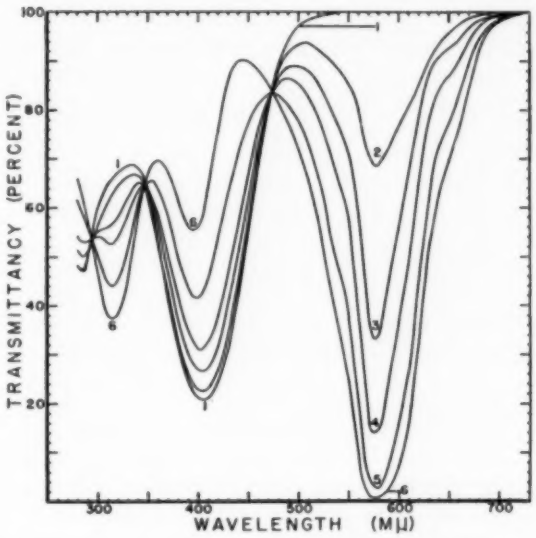


FIGURE 10. *Transmittancy curves for bromphthalein magenta E, 5×10^{-5} M in benzene, in 5-mm cell, with various concentrations of piperidine expressed as molar equivalents of the indicator.*

(1) None; (2) 0.1; (3) 0.3; (4) 0.5; (5) 1.0; (6) 10 (limiting curve).

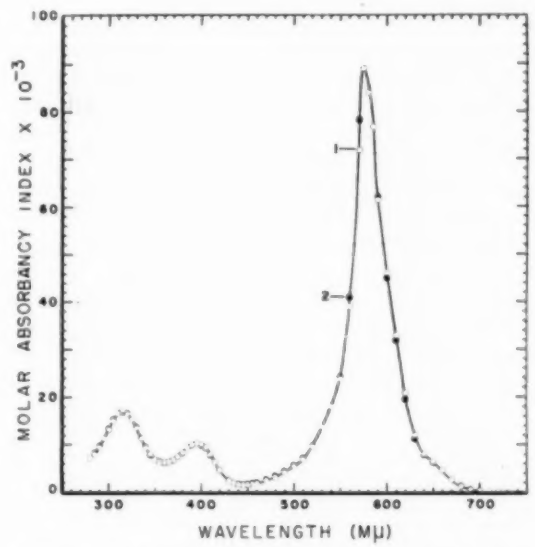


FIGURE 11. *Typical molar absorbancy curves for products of complete reaction of bromphthalein magenta E in benzene with secondary aliphatic amines.*

(1) Bromphthalein magenta E, 2.5×10^{-5} *M*, plus 40 molar equivalents of diethylamine; (2) bromphthalein magenta E, 5×10^{-5} *M*, plus 10 molar equivalents of piperidine.

The transmittancy curves for solutions containing bromphthalein magenta E with primary aliphatic amines were temporarily ignored in the discussion of figure 9. On first inspection, the primary amines seem to resemble the secondary amines in behavior, except in producing a smaller effect at a given concentration. The curves assembled in figure 9 were obtained for solutions that contained 4×10^{-5}-*M* BPM-E in a 1-cm cell, with 1.25 molar equivalent of the amine (5×10^{-5} *M*). In the order of increasing reactivity with bromphthalein magenta, the amines were: (2) Benzylamine; (3) *iso*butylamine; (4) *o*-aminodicyclohexyl; (5) *n*-propylamine; (6) isoamylamine; (7) *n*-butylamine; (8) *sec.*-butylamine; (9) *n*-amylamine; (10) *iso*propylamine; (11) cyclohexylamine; (12) *n*-octylamine; and (13) *n*-heptylamine. 2-Amino-*n*-octane was in the relative position (9a). It should be emphasized that quantitative relationships are only provisional, because the experiments were of an exploratory nature, and minor variations in temperature had a noticeable effect upon the values obtained.

Data for values of pK_b in water for some primary aliphatic amines are presented in table 4. The lack of agreement in the values of pK_b for a given amine are due not only to the varied nature of the experimental procedures used by different investigators but also to diversity in the assumptions made in calculations from the data. It is plainly evident, however, that the primary aliphatic amines exhibit lower basicity in water than the corresponding secondary amines. If the values cited first in table 4 are used as the basis for comparison, there appears to be a trend toward greater ionization for the homologous series of normal primary aliphatic amines from *n*-propylamine through *n*-heptylamine, and a reversal toward lower basicity with further increase in molecular weight. Other effects of structure upon the ionization constants of primary amines in water were noted by Bredig (footnote d, table 4). The data for the reaction of bromphthalein magenta with various primary amines in benzene show a smaller tendency toward reaction than that exhibited by secondary amines. Furthermore, as pK_b for water increases, there is seemingly a decrease in the reactivity in benzene. This is the reverse of the relationship found for five of the tertiary amines and arylguanidines. The data for the reactions of different primary amines with bromphthalein magenta in benzene are too provisional to warrant detailed speculation, but it seems likely that the reaction of the amines with water and with bromphthalein magenta may be obscured by lack of knowledge as to the real extent of hydration and ionization of these bases in water and of any self-association of the amine.

TABLE 4.—*Dissociation constants of some primary aliphatic amines in water* [a]

Compound	pK_b
Methylamine	[b] 3.36; [c] 3.40; [d], [e] 3.30; [f] 3.41
Ethylamine	[c] 3.34; [d] 3.13; [e] 3.25; [f] 3.40
(3) n-Propylamine	[c] 3.41; [d] 3.32; [e] 3.33
(5) n-Butylamine	[g], [b] 3.39
(7) n-Amylamine	[g] 3.37; [f] 3.33
n-Hexylamine	[g] 3.36
(11) n-Heptylamine	[g] 3.34
(10) n-Octylamine	[g] 3.35
n-Nonylamine	[g] 3.36
n-Decylamine	[g] 3.36
n-Undecylamine	[g] 3.37
n-Dodecylamine	[g] 3.37
n-Tridecylamine	[g] 3.37
n-Tetradecylamine	[g] 3.38
n-Pentadecylamine	[g] 3.39
(8) *Iso*propylamine	[c] 3.37; [e] 3.28
(2) *Iso*butylamine	[g] 3.57; [c] 3.59; [e] 3.51
(6) *sec*-Butylamine	[c] 3.44; [e] 3.34
(4) *Iso*amylamine	[c], [g] 3.40; [e] 3.30; [f] 3.33
(9) Cyclohexylamine	[f] 3.31; [f] 3.80
(1) Benzylamine	[b] 4.63; [c] 4.70; [e] 4.61

[a] In some cases, measurements were made in aqueous alcohol, and the value for aqueous solution was obtained by extrapolation. Values are for 25° C unless otherwise indicated. The numbers to the left of the names of the compounds indicate the apparent order of increasing reactivity with bromphthalein magenta *E* in benzene.

[b] H. S. Harned and B. B. Owen, J. Am. Chem. Soc. **52,** 5079 (1930).

[c] J. Kendall, Int. Crit. Tables **VI,** 260 (National Research Council, Washington, D. C., 1929).

[d] T. S. Moore and T. F. Winmill, J. Chem. Soc. (London) **101,** 1635 (1912).

[e] G. Bredig, Z. phys. Chem. **13,** 289 (1894).

[f] H. T. S. Britton and W. G. Williams, J. Chem. Soc. (London) 796 (1935); values are for 18° C.

[g] C. W. Hoerr, M. R. McCorkle, and A. W. Ralston, J. Am. Chem. Soc. **65,** 328 (1943); probable accuracy, ±0.01 unit.

[h] L. D. Goodhue and R. M. Hixon, J. Am. Chem. Soc. **56,** 1329 (1934).

[i] N. F. Hall and M. R. Sprinkle, J. Am. Chem. Soc. **54,** 3469 (1932).

[j] A. Waksmundzki, Roczniki Chem. **18**[1], 865 (1938), through Chem. Abstracts **33,** 6688 (1939); values are for 18° to 20° C.

[k] W. H. Carothers, C. F. Bickford, and G. J. Hurwitz, J. Am. Chem. Soc. **49,** 2908 (1927).

Figure 12 contains curves for the reaction of 5×10^{-5}-*M* bromphthalein magenta E with successively greater quantities of *n*-heptylamine in a 5-mm cell. The curves show the effects of the following concentrations of the amine, expressed as molar equivalents of the indicator: (1) none; (2) 1; (3) 2; (4) 5; (5) 20; (6) 80 (limiting curve). At the higher concentrations of amine, it becomes apparent that the main absorption band has two

maxima, one near 540 mμ and the other near 575 mμ. These are the wavelengths of maximum absorbancy for solutions of bromphthalein magenta with tertiary amines and secondary amines, respectively. In figure 12 three well-defined isosbestic points are evident in the spectral region 290 to 295 mμ, 345 to 350 mμ, and 465 to 470 mμ. Two single absorption bands occur near 315 mμ and 390 mμ, and transmission bands near 335 and 440 mμ. The ratio of the molar absorbancy indexes at 575 mμ and 540 mμ is not constant but falls from nearly 2.5 for one molar equivalent of base to approximately 1.2 for five or more molar equivalents. Dimerization of the base or formation of still larger aggregates at the higher concentrations may be a partial explanation for this behavior. The absorbancy curve is affected by fluctuations of only a few tenths of a degree centigrade, and when the temperature is lowered there is some evidence of a strengthening of the band near 575 mμ at the expense of the band near 540 mμ.

In figure 13 are shown typical molar absorbancy curves for the salts formed by the complete reaction of bromphthalein magenta with three different primary amines in benzene. These curves were obtained for 5×10^{-5}-M bromphthalein magenta E with the following amines: (1) Cyclohexylamine (100 molar equivalents); (2) *n*-heptylamine (80 molar equivalents); (3) *n*-octylamine (80 molar equivalents). The general pattern of the absorption bands is the same throughout the spectrum for the three salts, and the intensity of absorbancy is practically the same in the region 280 to 420 mμ. In the main band the total intensity differs for the three substances, and the ratios of the molar absorbancy indexes of the partial bands at 575 and 540 mμ are approximately 1.25, 1.17, and 1.14 for salts 1, 2, and 3, respectively. The differences do not appear to be due to instability, for solutions that contained up to 100 molar equivalents of any one of these bases gave constant transmittancy values for at least 4 hours.

5. Preparation and Absorption Spectra of Tetraalkylammonium Salts of Bromphthalein Magenta

Preparation.—To prepare the tetraethylammonium salt of bromphthalein magenta E, a solution of bromphthalein magenta E in benzene was shaken mechanically for 10 minutes with an ap-

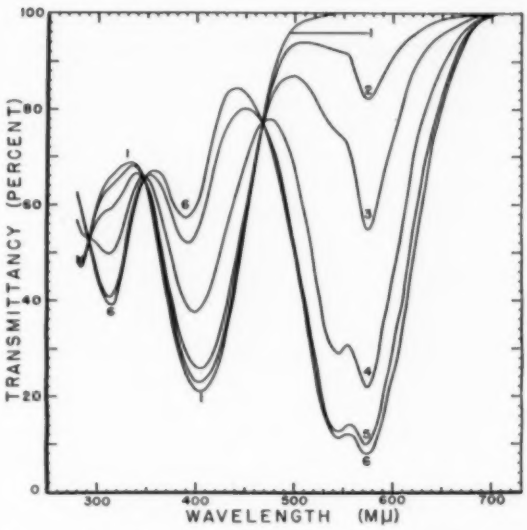


FIGURE 12. *Transmittancy curves for bromphthalein magenta E, 5×10^{-5} M in benzene, in a 5-mm cell, with various concentrations of n-heptylamine expressed as molar equivalents of the indicator.*

(1) None; (2) 1; (3) 2; (4) 5; (5) 20; (6) 80 (limiting curve).

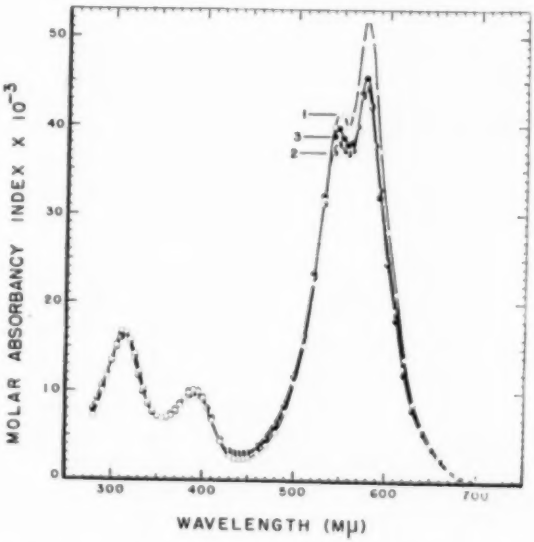


FIGURE 13. *Typical molar absorbancy curves for products of complete reaction of bromphthalein magenta E, 5×10^{-5} M in benzene, with primary aliphatic amines.*

Amine added and its concentration in molar equivalents of the indicator: (1) Cyclohexylamine, 100; (2) *n*-heptylamine, 80; (3) *n*-octylamine, 80.

proximately 2-percent aqueous solution of tetraethylammonium hydroxide. The salt precipitated as a shiny dark-green solid, which was filtered with suction after decantation of most of the benzene. The solid was extracted twice with a little hot benzene, the benzene was decanted, and the moist residue was dried at about 60° C. A benzene solution of the salt reacted with cellulose, as shown by the fact that when poured through dry filter paper the first portions of filtrate were green instead of the original blue. Moreover, a sample of the solid salt was decomposed to a gummy orange solid upon extraction with benzene in a Soxhlet apparatus. It is reasonable to suppose that this decomposition was caused by the strong tendency of quaternary ammonium bases, like other strong bases, to combine with and even dissolve cellulose [64]. The reaction may be visualized as an alcoholysis of the salt, with cellulose playing the role of an alcohol.

Analysis of purified salt: Calculated for $C_{30}H_{33}O_4NBr_4$: C, 45.5; H, 4.2; N, 1.8; Br, 40.4. Found: C, 45.2; H, 4.6; N, 1.6; Br, 40.6.

In the preparation of the tetra-*n*-butylammonium salt of bromphthalein magenta B, a benzene solution of bromphthalein magenta B was shaken mechanically with an aqueous solution of tetra-*n*-butylammonium hydroxide. The latter solution was the filtrate from the reaction of aqueous tetra-*n*-butylammonium iodide with an excess of freshly precipitated silver oxide. The tetrabutylammonium salt remained dissolved in the benzene layer, giving it a deep-blue color, but was precipitated upon the addition of petroleum ether. When the precipitate was redissolved in benzene and allowed to crystallize slowly at room temperature, large bronze-colored crystals were formed. *Analysis:* Calculated for $C_{40}H_{53}O_4NBr_4$: C, 51.6; H, 5.7; N, 1.5; Br, 34.3. Found: C, 51.5; H, 5.8; N, 1.4; Br, 34.5.

Absorption spectra.—Molar absorbancy curves for the tetraethylammonium salt of bromphthalein magenta E and for the tetra-*n*-butylammonium salt of bromphthalein magenta B in benzene are shown in figure 14. Curves 1B to 6B are for the butyl derivative, at the following successive molar concentrations: 10^{-4}, 4×10^{-5}, 2×10^{-5}, 10^{-5}, 5×10^{-6}, and 2×10^{-6}. The less concentrated solutions were prepared by dilution of the 10^{-4}-*M* stock solution. Curves 4E to 6E are for the ethyl derivative at the following successive molar concentrations: 10^{-5}, 5×10^{-6}, and 2×10^{-6}. The stock solution in this case was 10^{-5} *M*, which is about the limit of the solubility of the ethyl derivative in benzene. Both salts possess a strong absorption band near 600 to 605 mμ (16,670 to 16,530 cm^{-1}) and three shallow absorption bands in the ultraviolet region. However, the molar absorbancy at 605 mμ does not adhere to Beer's law but decreases with dilution, falling from the maximum absorbancy index 82,000 for 1B to the index 51,000 for 6B. The indexes at 605 mμ for 4E and 6E are, respectively, 62,500 and 42,000.

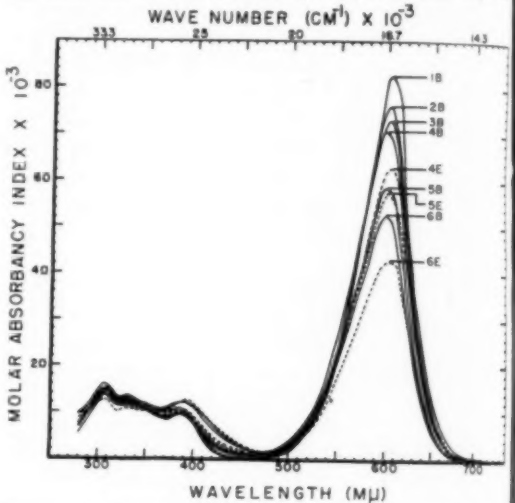


FIGURE 14. *Molar absorbancy curves for tetraalkylammonium salts of bromphthalein magenta in benzene at various concentrations.*

(1B to 6B) Tetra-*n*-butylammonium salt of bromphthalein magenta B; (4E to 6E) tetraethylammonium salt of bromphthalein magenta E. Concentrations: (1) 10^{-4} *M*; (2) 4×10^{-5} *M*; (3) 2×10^{-5} *M*; (4) 10^{-5} *M*; (5) 5×10^{-6} *M*; 6) 2×10^{-6} *M*.

On the other hand, the molar absorbancy near 400 mμ, the region in which bromphthalein magenta in the form of the free acid has a strong absorption band, shows a slight increase with dilution. The rate of change of the indexes with dilution is somewhat irregular.

In accounting for the change in the intensity of absorbancy with dilution, one possibility to consider is an equilibrium between ion pairs and quadrupoles, such as that which has been postulated to interpret conductance phenomena [65].[24]

[24] As no Tyndall beam was detected, and no precipitate was visible in any of the flasks, even after 10 months, the salts may be assumed to have been molecularly dispersed in the benzene solution. Kraus and coworkers in studying the conductance of tetraalkylammonium salts in anisole, ethylene dibromide, and ethylene dichloride observed dilution effects which they attributed to sorption [58].

It is questionable whether such an equilibrium would account for spectral changes of the kind observed, and also whether quadrupoles would exist in measurable quantity at such low concentrations of solute. A second possible explanation is found in evidence of the reaction of the salt with glass. The flasks that contained the solutions corresponding to curves 1B and 4E were drained, washed three times with benzene, and allowed to drain until dry. Freshly boiled distilled water containing phenolphthalein added to each flask produced a pink color that was more vivid in the flask that had contained the ethyl (less soluble) derivative.[25] A possible interpretation of this behavior is reaction of the tetraalkylammonium salt with $HO-Si(-O-)_3$ at the glass surface,[26] giving $R_4N^+-OSi(-O-)_3$ with bromphthalein magenta as byproduct, followed by hydrolysis of the R_4N^+-O-Si portion of the complex upon the addition of water, to yield R_4N^+, OH^-, and $HO-Si(-O-)_3$. The greater size of the hydrocarbon groups in the tetrabutylammonium salt not only explains its higher solubility in benzene, but also might well impede its reaction with glass.

In figure 15 the molar absorbancy curve for the tetraethylammonium salt of bromphthalein magenta *E* in benzene (curve 1) is compared with those obtained for solutions in ethyl acetate (curve 2), ethylene dichloride (curve 3), and 95-percent ethyl alcohol (curve 4). Curve 4 shows two well-marked absorption bands near 26,000 and 32,000 cm^{-1}, in place of the three shallow bands in the ultraviolet noted previously for the solution in benzene (page 242). In curve 3, a shift of about 2,500 cm^{-1} (10mμ) towards the infrared is shown for the main absorption band. In other respects the molar absorbancy curves obtained for the tetraethylammonium salt in the four solvents are qualitatively similar. The only marked difference is in the intensity of the main absorption band, which is not significantly different for the solutions in ethylene dichloride and 95-percent alcohol but is lower for the solution in ethyl acetate and still lower for that in benzene. The molar absorbancy curves for 5×10^{-5}-*M* solutions in ethylene dichloride and alcohol did not differ from those

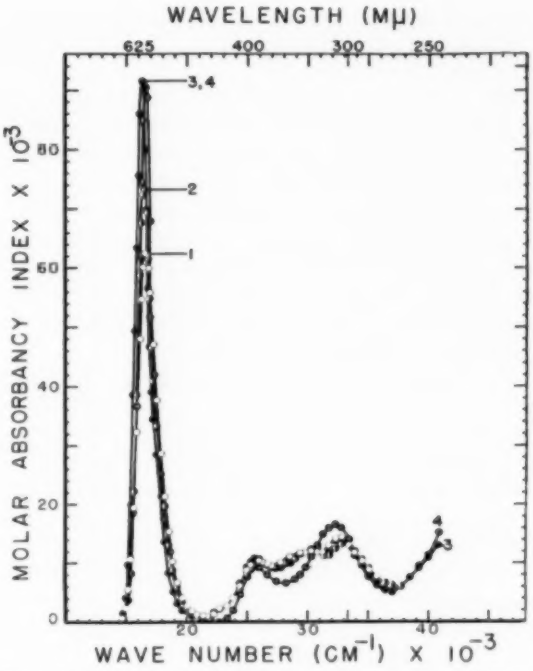


FIGURE 15. *Molar absorbancy curves for tetraethylammonium salt of bromphthalein magenta E.*

(1 to 3) 10^{-5} *M* in 1-cm cell, in benzene, ethyl acetate, and ethylene dichloride, respectively; (4) 5×10^{-5} *M*, 2.5-mm cell, in 95-percent ethyl alcohol.

obainted for 10^{-5}-*M* solutions. Furthermore, the same curve was obtained for the potassium salt of bromphthalein magenta E in alcohol as for the tetraethylammonium salt. Due to limited solubility, absorbancy curves for 5×10^{-5}-*M* solutions of the tetraethylammonium salt in benzene and ethyl acetate could not be obtained. It can be noted, however, that at twice this concentration the molar absorbancy at 605mμ found for the *tetrabutyl*ammonium salt in benzene (1B, fig. 14) is less than that found for the tetraethylammonium salt in ethylene chloride and in alcohol. As there is doubt that the ethyl and butyl salts remained entirely in solution in benzene (see preceding paragraph), the low intensity of the main absorption band for the solution in benzene may, at least tentatively, be accounted for on that basis. A similar explanation serves for the data found for the ethyl acetate solution. The dielectric constant for ethyl acetate is approximately 4.5, or about twice as great as that for benzene; this probably explains its somewhat higher solvent action upon the salts in question.

[25] Negative tests were obtained with three flasks chosen at random from the stock. Other organic bases have been observed to undergo adsorption or reaction at the surface of glassware.

[26] The oxygen atoms which are not attached to hydrogen are considered to be linked to silicon atoms in the glass network [66, 67].

6. Relative Stabilities of the Substituted Ammonium Salts of Bromphthalein Magenta in Benzene

The primary, secondary, tertiary, and quaternary ammonium salts of bromphthalein magenta in benzene appear to increase in stability in the order named. Complete quantitative data concerning the rate and extent of changes upon standing have not been obtained, because slight fluctuations in temperature are great enough in some cases to account for part of the observed changes. It is possible, however, to make several statements about the stability of the solutions.

In regard to solutions containing primary amines, it has already been stated that 5×10^{-5}-M bromphthalein magenta in benzene with up to 100 molar equivalents of (a) cyclohexylamine, (b) *n*-heptylamine, or (c) *n*-octylamine showed no measurable change for at least 4 hours. In 3 weeks solution (a) had lost all its color. With only 1 molar equivalent of amine, the solution was stable for at least a day and retained some color at the end of 3 weeks. The same concentration of dye with 180 molar equivalents of *n*-butylamine showed noticeable fading in two hours—a 9-percent decrease in transmittancy at 540 mμ. When kept for several weeks all of the solutions showed fading, which was greater at the higher concentrations of base. The most concentrated solutions when almost colorless contained minute quantities of iridescent green crystals that gave a blue solution in alcohol.

A solution that contained 1.4 molar equivalent of piperidine showed changes after 2 days that appeared to involve both fading and reaction of the base with glass. There are two positions in the dye molecule where reaction with a primary or secondary amine might occur. One is the ester group, that might be saponified if the solution absorbed water. The other position is the central carbon, which can add RNH— or R_2N—, with simultaneous addition of H— to the quinoid oxygen atom. This latter reaction is analogous to the conversion of triphenylmethane dyes to the carbinol form in aqueous solutions of relatively high pH. An example is the fading of some of the sulfonephthalein indicators in alkaline media. A smaller tendency in this direction can be expected for primary amines that contain a bulky R group, and secondary amines should react less readily than primary amines. The amines that have been investigated behave in accordance with this prediction. Thus it happens that the amines showing the least stability with bromphthalein magenta in benzene are those most easily determined in water; that is, primary amines of low molecular weight. The bases with large alkyl groups are difficultly soluble in water, but give sufficiently stable solutions with bromphthalein magenta for measurement of their reactivity with this compound.

Tertiary amines cannot react with the central carbon atom in the manner discussed in the preceding paragraph. It is found experimentally that the transmittancy of solutions of bromphthalein magenta with tertiary amines or diarylguanidines changes by 1 or 2 percent at most in a month, and even after a year there is no fading evident to the eye.

As for the tetraalkylammonium salts of bromphthalein magenta in benzene, changes such as those discussed in section IV–5 were not evident for solutions that were not diluted. The tetraethylammonium salt gave the same transmittancy values after 9 days, and showed no visual change in over a year. Apparently the extent of change is limited by the area of the glass surface with which the solution comes in contact.

It might be pointed out that as the indicator alone and many of the amines alone give stable solutions in benzene, it has been possible to keep stock solutions of both indefinitely. The desired mixtures can then be prepared as needed.

7. Relative Strengths of Some Organic Bases in Benzene

In the preceding sections, several statements were made about the comparative strengths of organic bases as evidenced by their ionization constants in water and their reactivity with bromphthalein magenta in benzene. Some consideration was also given to the probable mechanism of the reaction of the bases with the indicator acid. Before discussing the reaction constants for organic bases, a few words should be said about the reactions from a purely empirical standpoint. In figure 16 the number of moles of base per mole of indicator acid has been plotted against the absorbancy at 540 mμ [27] for 5×10^{-5}-M bromphthalein

[27] Values for absorbancy, which is the negative logarithm of the transmittancy, can be read directly from the instrument but not so precisely as the values for transmittancy. The absorbancies given in figure 16 were computed from the transmittancies. Absorbancy is preferred to the equivalent term "optical density."

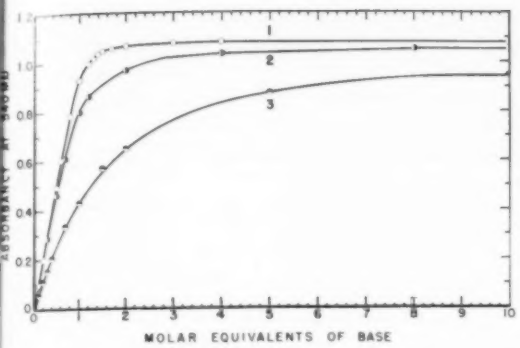


FIGURE 16. *Change in absorbancy produced by adding different amounts of base, expressed in molar equivalents of the indicator, to 5×10^{-5}-M bromphthalein magenta in benzene.*

(1) Bromphthalein magenta E plus di-o-tolylguanidine; (2) bromphthalein magenta E plus diphenylguanidine; (3) bromphthalein magenta B plus triethylamine.

magenta in benzene with (1) di-*o*-tolylguanidine; (2) diphenylguanidine; and (3) triethylamine. The smoothness of the curves makes it possible to obtain by interpolation the absorbancy for other concentrations of base. In fact, a good approximation of the correct curve may be drawn from only a few experimental points. Equally smooth curves were obtained for the reaction of bromphthalein magenta with secondary amines. It is apparent from the figure that the reactivity of the three bases decreases in the order given. The reaction had reached completion with 10 molar equivalents of di-*o*-tolylguanidine. Slightly more than this quantity of diphenylguanidine was required, and about ten times as much triethylamine. In the reaction of any one of these bases with bromphthalein magenta, intermediate color tones can be detected visually with great ease. The possibility of application to rapid colorimetric analysis is therefore evident.

In section IV–4, a simple mechanism for the reaction of bromphthalein magenta with organic bases in benzene is proposed, namely, the addition of the indicator acid to the base in a 1:1 ratio. This may be an oversimplification, or may not be applicable in all cases. However, it serves as a starting point for numerical comparisons. The type of equation for the reaction was $B+HOR \rightleftarrows \overset{+}{B}H\overset{-}{O}R$ or, using the symbols A and S to represent the acid HOR, and the product, $\overset{+}{B}H\overset{-}{O}R$, respectively, the equilibrium constant may be formulated as follows:

$$[S]/\{[B]\times[A]\}=K_b. \qquad (18)$$

From this, the following expression is obtained:

$$\log\{[S]/[A]\}-\log[B]=\log K_b. \qquad (19)$$

When the value of [S]/[A] is unity, $\log K_b=-\log[B]$. In figure 17 experimental data for the reaction of bromphthalein magenta with five different bases are presented, with values for $\log\{[S]/[A]\}$ as ordinates and values for $-\log[B]$ as abscissas. In order to calculate the values for [S], [A], and [B], data such as those used in plotting figure 16 were used. The ratio of the absorbancy for a given concentration of base to the absorbancy for complete conversion of the indicator acid to the salt gives the fraction of the indicator that was converted to the salt. This fraction multiplied by the molar concentration, C_a, of indicator originally in the solution gives [S], the concentration of the salt in moles per liter. The concentra-

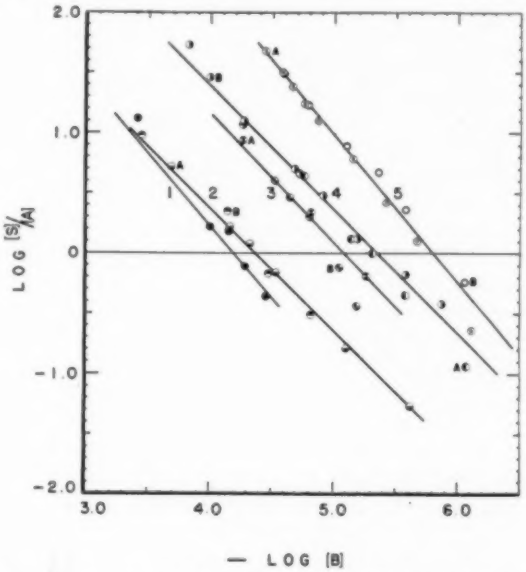


FIGURE 17. *Graphical evaluation of equilibrium constants for the reaction of bromphthalein magenta with different bases in benzene.*

(1) Bromphthalein magenta E, 1.6×10^{-4} *M*, with di-*n*-butylamine; (2A) 5×10^{-5}-*M* bromphthalein magenta B, and (2B) 4×10^{-5}-*M* bromphthalein magenta E, with triethylamine; (3A and 3B) 5×10^{-5}-*M* bromphthalein magenta E with piperidine; (4A) 4×10^{-5}-*M* bromphthalein magenta E, and (4B) 5×10^{-5}-*M* bromphthalein magenta B, wish diphenylguanidine; (5A) 5×10^{-5}-*M* bromphthalein magenta E, and (5B) 2.5×10^{-5}-*M* bromphthalein magenta B, with di-o-tolylguanidine.

tion of residual acid, [A], is $C_a - [S]$. If C_b is the molar concentration of base added to the mixture, the concentration of base in the equilibrium mixture, [B], equals $C_b - [S]$. From a series of values for [S], [A], and [B] calculated in this way for solutions containing the same initial concentration of the indicator acid with different amounts of base, the curves shown in figure 17 were constructed.[28] These curves correspond to the following mixtures in benzene as solvent: (1) Bromphthalein magenta E, 1.6×10^{-5} *M*, with di-*n*-butylamine; (2A) bromphthalein magenta B, 5×10^{-5} *M*, with triethylamine; (2B) bromphthalein magenta E, 4×10^{-5} *M*, with triethylamine; (3A and 3B) bromphthalein magenta E, 5×10^{-5} *M*, with piperidine; (4A) bromphthalein magenta E, 4×10^{-5} *M*, with diphenylguanidine; (4B) bromphthalein magenta E, 5×10^{-5} *M*, with diphenylguanidine; (5A) bromphthalein magenta E, 5×10^{-5} *M*, with di-*o*-tolylguanidine; and (5B) bromphthalein magenta B, 2.5×10^{-5} *M*, with di-*o*-tolylguanidine.

The relationship postulated in eq 2, if valid, should give linear curves, with unit negative slope. The lines were drawn so as to give greatest weight to the values that can be measured with the greatest precision, that is, those for which the ratio [S]/[A] is near unity. The intercepts on the *x*-axis at zero value for log {[S]/[A]} give the values for log K_b. These values, together with the values for the slope, are given in table 5. It is apparent from figure 17 that if the lines had been drawn so as to give slightly different values for the slope, the values obtained for log K_b would have been practically unchanged. For triethylamine and diphenylguanidine, the slope of the line differs by only 3 percent from the theoretical value. The greatest deviation, 23 percent, occurs in the case of ditolylguanidine. Factors such as polymerization of the base may complicate the reaction. For example, Hunter and Marriott [68], from measurements of molecular weights in naphthalene, concluded that 1,3-diphenylguanidine and related compounds are partly associated in this solvent. However, the solutions used in the experiments described in this paper were so dilute that the degree of association of the base should be very small. The authors are inclined to attribute discrepancies largely to experimental difficulties, such as minor fluctuations in temperature, the tendency of the base to react or become absorbed at a glass surface, and the unavailability of adsorption cells of suitable length at the time the spectrophotometric measurements were made.

[28] To be exact, the values used in constructing the curves were the averages from two sets of measurements—one for the wave length 400 or 405 mμ and the other for the wavelength 540 or 580 mμ.

TABLE 5.—*Values of log K_b, K_b, and slope for reaction of organic bases with bromphthalein magenta in benzene*

Organic base	log K_b	K_b	Slope of curve
Di-n-butylamine	4.19	$1.5_9 \times 10^{4}$	−1.07
Triethylamine	4.36	2.3×10^{4}	−1.01
Piperidine	5.08	1.2×10^{5}	−1.08
Diphenylguanidine	5.35	2.2×10^{5}	−1.01
Ditolylguanidine	5.80	6.3×10^{5}	−1.23

In figure 18 the numerical values of log K_b in benzene for the five bases are plotted as ordinates, with the values of pK_b in water as abscissas. It will be remembered that pK_b is the reciprocal of the logarithm (negative logarithm), of the ionization constant of the base in water (see section IV–4). From the figure it is evident that the relationship between log K_b in benzene and pK_b in water is remarkably constant for triethylamine and the two arylguanidines. Whether a consistent relationship holds for other tertiary amines and secondary amines can be determined only by further studies. We hope to extend these studies, under improved experimental conditions.

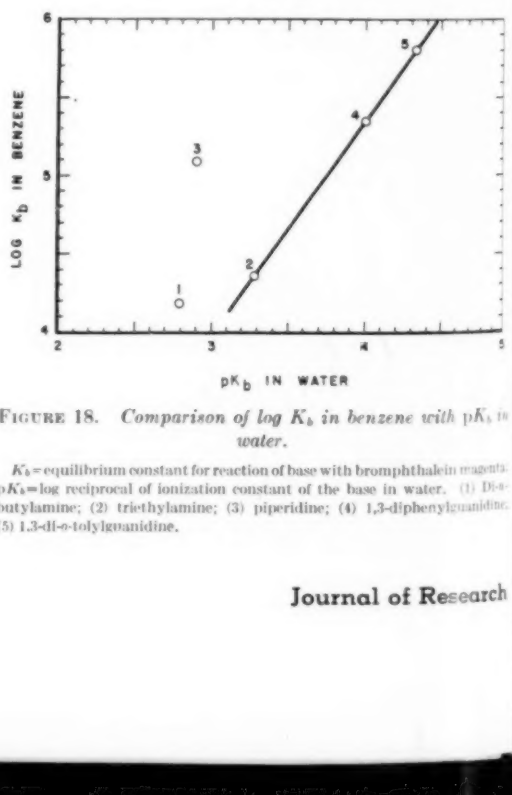


FIGURE 18. *Comparison of log K_b in benzene with pK_b in water.*

K_b = equilibrium constant for reaction of base with bromphthalein magenta; pK_b = log reciprocal of ionization constant of the base in water. (1) Di-n-butylamine; (2) triethylamine; (3) piperidine; (4) 1,3-diphenylguanidine; (5) 1,3-di-o-tolylguanidine.

8. Explanation for the Different Color Reactions of Bromphthalein Magenta With Organic Bases in Benzene

As described earlier in this paper, the color change accompanying the reaction of an organic base with bromphthalein magenta in benzene is different for primary, secondary, and tertiary amines. Moreover, the quaternary ammonium salts, while resembling most nearly the product with secondary amines, show characteristic differences from the partially substituted ammonium salts of the indicator.

The conductance of quaternary ammonium salts in benzene and other solvents of low dielectric constant has been studied extensively, particularly by Walden and Kraus and their coworkers [28, 57b]. These studies have given strong experimental support to Bjerrum's theory of ion-pairs [69]. In harmony with the views of these authors, the tetraethylammonium and tetra-*n*-butylammonium salts of bromphthalein magenta may be considered to exist in dilute benzene solution predominantly as individual ion-pairs, M^+A^-, M^+ representing the cation, and A^- the anion. To a minor degree, dissociation of M^+A^- into M^+ and A^-, or association into triple-ions, MAM^+ and AMA^-, or into still more complex groupings may occur. According to Bjerrum, the extent of dissociation or association will not be evident in the optical behavior of the salt.

The difference in the optical behavior of the quaternary ammonium (blue) and tertiary ammonium (magenta) salts of bromphthalein magenta in benzene is paralleled by differences in conductances found for salts of the two classes [28, 57]. In the tertiary ammonium salts, the existence of a proton in place of one of the alkyl groups of the quaternary ammonium salt permits a closer approach of the positive and negative ions to each other, with a consequent increase in the Coulombic force. However, the change in behavior that accompanies this modification in structure is undoubtedly due in part to the formation of a hydrogen bridge between the nitrogen of the base and the oxygen of the indicator acid.[29] The structure of the salt can be indicated in a general way by the formula $B{-}H^+ \ldots \bar{O}R$.

In accounting for the purple-blue color of benzene solutions of the secondary ammonium salts of bromphthalein magenta, other observations are of value. During titrations of secondary amines in benzene with bromphthalein magenta as indicator, it was discovered that the initial effect of adding an acid to the purple-blue solution was to produce a magenta color, which was changed to yellow by an excess of acid. When an excess of base was then added, the magenta color reappeared, but the purple-blue color could not be restored by any quantity of base.[30] Further tests showed that the conversion of the purple-blue color to magenta was not characteristic of acids but was produced by a variety of oxygen compounds, such as dioxane, acetone, ethyl acetate, and alcohols.[31] The magenta color is very similar to that of mixtures of bromphthalein magenta with tertiary amines in benzene. Spectrophotometric measurements of solutions that contained 5×10^{-5}-*M* bromphthalein magenta E and 5×10^{-4}-*M* piperidine in benzene, with from 4 to 6 percent of absolute ethyl alcohol by volume, showed that the absorbancy curves of these solutions are very similar to, though not identical with, the characteristic limiting curve for the reaction of a tertiary amine with bromphthalein magenta in benzene.[32]

It can hardly be questioned that the change in color produced by dioxane, acetone, and other oxygen compounds arises from their interaction with the amine. Numerous studies of hydrogen bridges between amines and other types of compounds have been made by infrared and Raman spectroscopy, solubility measurements, and other methods. From these studies, the ability of amines to form hydrogen bridges with compounds containing active hydrogen, such as alcohols and haloforms, appears to be well established. Associated molecules in which the hydrogen of the amine is bridged to the oxygen of another molecule seem to be less common. However, Gordy [70] obtained spectroscopic evidence for association of this type in mixtures of aniline with acetone,

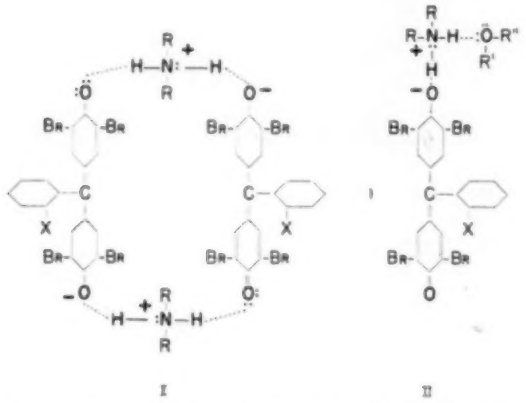
[29] See discussion in the introduction and in section IV–4.

[30] It was noted subsequently that a similar behavior has been described for bromphenol blue, the sulfonephthalein analog of bromphthalein magenta [41].

[31] Normal mercaptans with 8, 10, 12, or 14 carbon atoms were without effect. For additional compounds that produce the effect, see section IV–10.

[32] Continued addition of alcohol produced a gradual shift to the blue color and molar absorbancy curve characteristic of the tetraethylammonium or potassium salts of bromphthalein magenta in 95-percent alcohol. In the absence of piperidine, the effect of 4 to 6 percent of alcohol on the absorbancy curve of the indicator was almost negligible. The molar absorbancy curve for 5×10^{-5}-*M* bromphthalein magenta in absolute ethyl alcohol is the same as that found for the potassium and tetraethylammonium salts in 95-percent alcohol, but in more concentrated solutions the presence of unchanged indicator is evident.

esters, and ethers. Furthermore, according to Huggins [15], *ammonium* hydrogens can be expected to form stronger bridges to oxygen compounds than *amino* hydrogens. Consequently, formula II in figure 19 provides a reasonable structure for the magenta complex formed in



I II

FIGURE 19. *Formulas postulated for dialkylammonium salts of bromphthalein magenta.*

I, Dimer formed in benzene and other inert solvents; II, monomer formed in presence of an oxygen compound. R and R' = alkyl groups; R'' = hydrogen, alkyl, or other organic radical.

benzene solutions of bromphthalein magenta, a secondary amine, and an oxygen compound, R′OR″. When the purple-blue salt is formed by reaction of bromphthalein magenta with a secondary amine in benzene, it is logical to assume that in the absence of an oxygen compound the hydrogen atom linked to oxygen in formula II is involved in some other type of bridge. A structure for the purple-blue salt is suggested in figure 19, formula I. That is, a dimer, stabilized by resonance, is postulated. This hypothesis is consistent with the fact that the color of the secondary ammonium salt of bromphthalein magenta is intermediate between that of the tertiary ammonium salt and the quaternary ammonium salt. Additional support is furnished by studies of the molar polarization of partially substituted ammonium picrates in benzene as a function of concentration, performed in this laboratory by A. A. Maryott. Tertiary ammonium picrates exhibit no association, while secondary ammonium picrates are associated appreciably in a way that results in some cancellation of the dipole moments of the individual molecules of the salt. Primary ammonium picrates are still more highly associated. When dioxane is used as solvent, there is no evidence of association for any one of the three types of salt [71].

As for the primary ammonium salts of bromphthalein magenta, which show optical resemblances to both the secondary ammonium and the tertiary ammonium salts, the following explanation is suggested. A dimer may form, as in the case of the secondary ammonium salt, but the third hydrogen of the ammonium ion can participate in bridge formation. As a result, the other nitrogen bridges are weakened.

It seems quite likely that phenomena of the kind so readily observed in the reaction of bromphthalein magenta with bases in benzene play an important role in numerous other acid-base reactions.

9. Reaction of Bromphthalein Magenta in Benzene With Miscellaneous Other Organic Bases

Ammonia.—A blue color results when dry ammonia is passed into bromphthalein magenta in benzene, and there is spectrophotometric evidence of an absorption band near 575 mμ. However, the product separates so rapidly from solution that quantitative studies are not possible.

Alkanolamines and other oxygen-containing amines.—As described in section IV–8, addition of an alcohol or other oxygen compound to the blue solution of bromphthalein magenta and a secondary amine in benzene changes the color to magenta. It is not surprising, therefore, that oxygen within the secondary amine molecule, as in morpholine or methylmonoethanolamine, $CH_3NHCH_2CH_2OH$, produces a similar effect. For example, a small amount of morpholine with bromphthalein magenta in benzene produces a blue color, which changes to magenta with continued addition of morpholine. A trace of methylmonoethanolamine produces a blue color slightly tinged with red, and a larger quantity gives a magenta color. Such color phenomena, which at first seemed puzzling, can readily be understood and predicted by applying the explanation offered in section IV–8. That is, when relatively few molecules of the oxygenated amine are present their *amino* groups react with the indicator acid to give the blue dimer; when surplus molecules are present their *oxygen* groups become linked through hydrogen with the nitrogen atoms in the dimer, and the dimer changes to the magenta-colored monomeric

salt. As the second reaction may occur before the first change is complete, there is likely to be evidence of the magenta color, together with the blue. The "normal" behavior of primary amines is likewise affected by oxygen groups within the molecule. Triethanolamine and *N*-methyl- or *N*-ethylmorpholine, on the other hand, give the magenta color produced by any tertiary aliphatic amine. The change in the color of bromphthalein magenta caused by any particular amine depends upon the character of its nitrogen atom (primary, secondary, or tertiary), upon the nature and number of its oxygen groups, and upon the concentrations of the amine and the indicator acid. The behavior of concentrated solutions in general is complicated by overlapping effects. Sensitive color reactions, such as those of bromphthalein magenta with simple amines in benzene, are helpful in gaining an understanding of more complex systems.

Alkaloids.—Alkaloids and other drugs of basicity comparable to that of aliphatic amines should show color reactions with bromphthalein magenta like those already described, provided steric or other factors do not interfere. A few tests with alkaloids containing tertiary amino groups gave results of the character anticipated. For example, a minute quantity of brucine, codeine, nicotine, quinine, or strychnine produced a magenta tone with bromphthalein magenta in benzene. Brucine, codeine, and nicotine were added as the free bases. The other two alkaloids, available as the sulfate, were dissolved in water, a trace of alkali was added, and the base was extracted with benzene. The benzene extract gave an immediate color reaction with the indicator.

One may predict the behavior of other alkaloids and drugs. For instance, coniine (α-*n*-propylpiperidine) should give a blue color with bromphthalein magenta in benzene, but the majority of the alkaloids are likely to give a magenta tone.

Aromatic amines.—Aniline, dimethylaniline, and similar aromatic bases are too weakly basic to show measurable reaction with bromphthalein magenta at concentrations such as those dealt with in the preceding sections of this paper. In amounts greater than about 100 molar equivalents, aniline or dimethylaniline produce a perceptible change of color from yellow to amber. With increasing amounts of the aromatic amine, the color gradually deepens. Spectrophotometric measurements showed that no stoichiometric reaction occurs.

Tetrahydroquinoline and decahydroquinoline.—Decahydroquinoline, even in a very small quantity, gives a blue color with bromphthalein magenta in benzene, showing the usual behavior of a secondary amine. The basicity of tetrahydroquinoline, on the other hand, is weakened by the partial aromatic character of the molecule. In optical behavior, it shows some similarity to primary aliphatic amines. Absorbancy curves for different mixtures of bromphthalein magenta E with tetrahydroquinoline in benzene are shown in figure 20. These curves correspond to 10^{-4}-*M* dye in a 2.5-mm cell, with the following concentrations of tetrahydroquinoline, expressed as molar equivalents of the indicator: (1) None; (2) 80; (3) 500; (4) 800; (5) 1,200; (6) 1,680; (7) 10,000. In addition to the absorption band at 575 mμ, there is evidence of a second band near 540 mμ. For concentrations of tetrahydroquinoline up to 800 molar equivalents (curve 4), there are two fairly sharp isosbestic points, one near 350 mμ and the other near 455 mμ. The absorption band

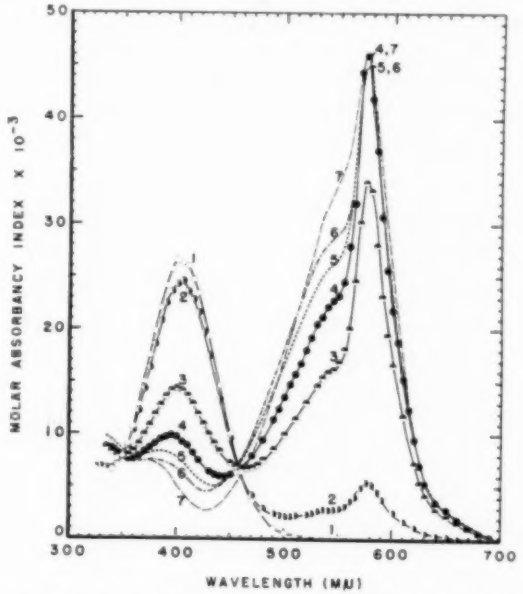


FIGURE 20. *Absorbancy curves for bromphthalein magenta E, 10^{-4}-M in benzene, 2.5-mm cell, with various concentrations of tetrahydroquinoline expressed as molar equivalents of the indicator.*

(1) None; (2) 80; (3) 500; (4) 800; (5) 1,200; (6) 1,680; (7) 10,000.

near 575 mμ reached a maximum index of 45,000 to 46,000 for about 800 molar equivalents of tetrahydroquinoline. When larger quantities of the base were added to the dye, the molar absorbancy index at 575 mμ remained fairly constant, but the absorbancy near 540 mμ increased, and the solution exhibited instability. Consequently, it was not possible to determine the limiting curve for the reaction. The lack of constancy of the molar absorbancy index near 575 mμ may be due to incipient precipitation of the salt. Tetrahydroquinoline was present in the comparison cell when the data for curves 2, 3, and 4 were obtained. Insufficient tetrahydroquinoline was available to permit its use in the comparison cell during the measurements for curves 5, 6, and 7. Furthermore, the two sets of curves were obtained with two different samples which contained different amounts of a brown impurity. The lack of constancy of the isosbestic point near 455 mμ in the case of curves 5, 6, and 7 is probably caused partly by the colored impurity and partly by incipient precipitation of the salt.

Alkyl-substituted guanidines.—This interesting class of nitrogen bases has been ignored by most investigators. A valuable contribution was made by Davis and Elderfield, who studied the effect upon the basicity of guanidine and its derivatives of variations in the kind of substituent and the position of substitution [72]. They pointed out that guanidine and its monoalkyl, *N,N*-dialkyl, and *N, N′, N″*-trialkyl derivatives give aqueous solutions comparable in basic strength with aqueous potassium hydroxide. On the other hand, monophenylguanidine is considerably less basic than monoalkylguanidines and *N, N′*-dialkylguanidines are weaker still. Pauling has suggested that resonance in the guanidinium ion may explain the high basicity of guanidine and the effect of substituents [17]. The study of acid-base reactions in organic media has been hampered by the lack of an organic base sufficiently strong to react with acetic, benzoic, and other weak organic acids. The data presented in section IV–7, indicate that *N, N′*-diphenyl- and di-*o*-tolylguanidine are more strongly basic in benzene than piperidine, which is generally considered to be the strongest of the aliphatic amines. *N, N′*-Dialkylguanidines should be more basic than the corresponding diarylguanidines, and *N, N′, N″*-trialkylguanidines, still stronger. Unfortunately, such compounds are not available commercially at the present time. However, interest in guanidine derivatives is growing, and it is hoped that eventually a series of symmetrical di- and trialkyl derivatives will be obtainable. One would expect that *N, N′, N″*-trialkylguanidines in which the alkyl groups contain 4 or more carbon atoms would possess sufficiently great basicity and solubility in inert solvents to prove useful in titrations of weak organic acids.

Experimental samples of *N,N′*-dicyclohexylguanidine hydrochloride, monododecylguanidine, and the salts of a few other monosubstituted guanidine derivatives were obtained from a commercial source.[33] Dicyclohexylguanidine hydrochloride was converted into the free base. As anticipated, this is more basic than diphenyl- or di-*o*-tolylguanidine. It also differs in giving a blue color instead of magenta when mixed with bromphthalein magenta in benzene. The usefulness of dicyclohexylguanidine is limited by its rather low solubility, which is little more than 0.0025 mole per liter of benzene. Monododecylguanidine also turns bromphthalein magenta blue and is difficultly soluble in benzene.

In other experiments, a suspension of monophenyl-, α-naphthyl-, or *p*-tolylguanidine carbonate in distilled water was shaken for a few moments with benzene. Partial hydrolysis of the salt occurred, and a sufficient quantity of the free guanidine derivative dissolved in the benzene layer to give a blue color upon the addition of bromphthalein magenta. Commercial methylguanidine sulfate behaved similarly when a trace of alkali was added to its aqueous solution and the latter was then shaken with benzene.

In figure 21 are shown the molar absorbancy curve for bromphthalein magenta in benzene (curve A) and the limiting curves for its reaction with the following organic bases: (1) Cyclohexylamine; (2) diethylamine; (3) triethylamine; (4) tetra-*n*-butylammonium hydroxide; (5) 1,3-dicyclohexylguanidine; (6) tetrahydroquinoline.[34] Curves 2, 3, and 4 show single maxima in the visible region. These occur at the following respective wave numbers ($cm^{-1} \times 10^{-3}$): 17.4, 18.4, 16.5. For curve A, the maximum in the

[33] The authors are indebted to the American Cyanamid Co. for these samples.

[34] The maximum molar absorbancy index for curve 4 is probably too low (see section IV–5). Curve 6 was calculated from the data shown as absorbancy curve 4 in fig. 20.

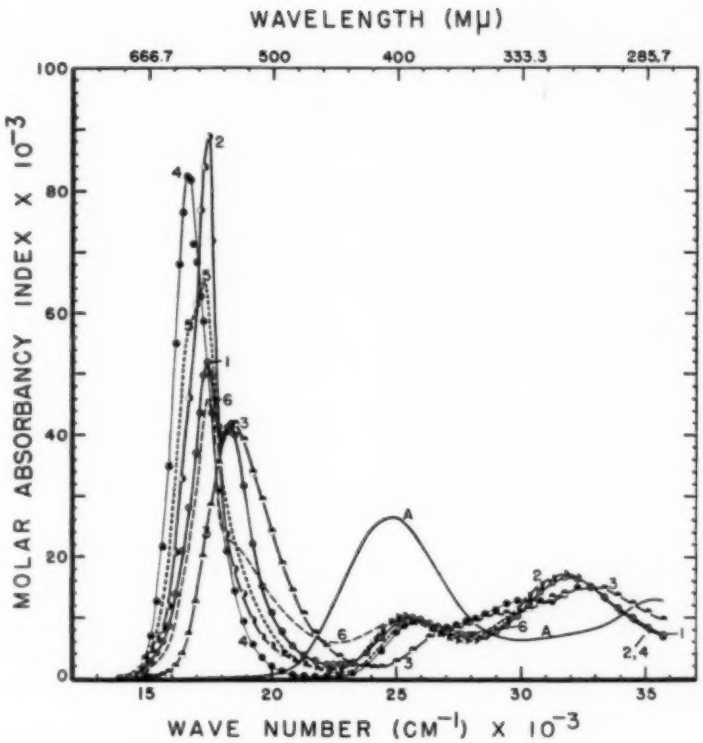


FIGURE 21. *Molar absorbancy curves for bromphthalein magenta and its salts with different organic bases in benzene.*

(A) Bromphthalein magenta E; (1) bromphthalein magenta E plus cyclohexylamine; (2) bromphthalein magenta E plus diethylamine; (3) bromphthalein magenta B plus triethylamine; (4) tetra-*n*-butylammonium salt of bromphthalein magenta B, $10^{-4}M$; (5) bromphthalein magenta E plus 1,3-dicyclohexylguanidine; (6) bromphthalein magenta E, $10^{-4}M$, plus 800 molar equivalents of tetrahydroquinoline.

visible region is located practically midway between the wave numbers 25.0 and 24.7 $cm^{-1} \times 10^{-3}$. In the other curves, there is evidence of two maxima in the visible spectrum, as follows: for curves 1 and 6, at 17.4 and 18.4; for curve 6, at 16.5 and 17.4. An explanation has been suggested (section IV–8) for the behavior of the primary, secondary, tertiary, and quaternary alkylammonium compounds (curves 1 to 4). The behavior of tetrahydroquinoline (curve 6) indicates that its salt with bromphthalein magenta is partly dimerized and partly in the monomeric form, while the curve for dicyclohexylguanidine suggests that its salt may exist partly as a dimer and partly as ion-pairs. However, because the dicyclohexylguanidine was only 98 to 99 percent pure, further experimentation is necessary before this interpretation can be accepted.

Trimethylamine oxide.—Although this paper is concerned primarily with bases in which the nitrogen atom acts as proton acceptor, it is of interest to compare the behavior of a tertiary amine oxide toward bromphthalein magenta in benzene with that of the tertiary amines, already discussed. Nylén found six amine oxides, including trimethylamine oxide, to be weaker bases than the corresponding amines in aqueous solution [73]. Some other published data for trimethylamine and trimethylamine oxide are summarized in table 6. Obviously the addition of an oxygen atom to the amine nitrogen changes appreciably the character of the base. Not only is the polar character of the molecule greatly enhanced and the carbon to nitrogen bond increased in length, but the picrate is much less dissociated in nitrobenzene. Consequently, one would predict stable salt formation from the reaction of trimethylamine oxide and bromphthalein magenta. The hygroscopicity of trimethylamine oxide is so great that preparation of a benzene solution of known concentration was

not attempted. However, some distinctive characteristics of the amine oxide could be detected.

TABLE 6.—*Comparison of trimethylamine and trimethylamine oxide*

Physical property	Trimethylamine	Trimethylamine oxide
Dipole moment (*D*)	[a] 0.65	[b] 5.02
Length of C—N bond (*A*)	[c] 1.47 ±0.02	[d] 1.54 ±0.02
Dissociation constant of picrate in nitrobenzene [e]	1.46×10^{-4}	1.73×10^{-4}

[a] L. G. Groves and S. Sugden, J. Chem. Soc. (London) 1779 (1937); gas.
[b] E. P. Linton, J. Am. Chem. Soc. **62**, 1945 (1940); in benzene at 45° C.
[c] L. O. Brockway and H. O. Jenkins, J. Am. Chem. Soc. **58**, 2036 (1936).
[d] M. W. Lister and L. E. Sutton, Trans. Faraday Soc. **35**, 495 (1939).
[e] C. A. Kraus, J. Phys. Chem. **43**, 231 (1939).

Trimethylamine was oxidized to trimethylamine oxide by 3-percent hydrogen peroxide, according to the method of Dunstan and Goulding [74], and crystallized as the dihydrate. The anhydrous product was obtained as a sublimate by Meisenheimer's procedure, namely, heating of the dihydrate in vacuo over phosphorus pentoxide [75]. Part of the sublimate was transferred quickly to a volumetric flask containing 5×10^{-5}-*M* bromphthalein magenta in benzene. Although the amine oxide did not dissolve completely, the solution immediately assumed a magenta tone. The absorbancy curve for this solution (curve 2) is shown in figure 22, together with the curves for

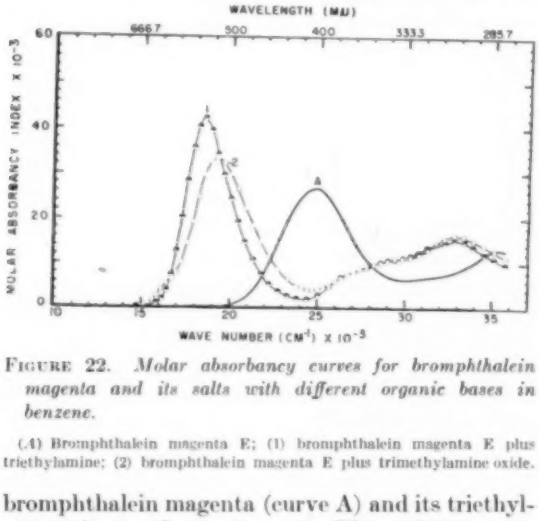


FIGURE 22. *Molar absorbancy curves for bromphthalein magenta and its salts with different organic bases in benzene.*

(A) Bromphthalein magenta E; (1) bromphthalein magenta E plus triethylamine; (2) bromphthalein magenta E plus trimethylamine oxide.

bromphthalein magenta (curve A) and its triethylammonium salt (curve 1). The solution was probably saturated with trimethylamine oxide, since the absorption spectrum did not change after further shaking and the lapse of 2½ days. As compared with the triethylammonium salt, the absorbancy curve for the salt of trimethylamine oxide is shifted approximately 20 mμ (700 cm^{-1}) toward shorter wavelengths, and the molar absorbancy index is 33,000 instead of 42,000. In the region 310 to 380 mμ (31,700 to 26,300 cm^{-1}) the absorbancy curves for the triethylammonium and trimethylamine oxide salts are nearly identical. Curve 2 intersects curve A at the approximate wavelength 450 mμ (22,220 cm^{-1}), while curve 1 intersects curve A near 460 mμ (21,740 cm^{-1}). Degeneracy of the resonance structures of the anion of bromphthalein magenta appears to be less complete in the salt formed by combination of the indicator with trimethylamine oxide than in the salts formed by its combination with trialkylamines, indicating that the proton of the indicator acid undergoes a smaller displacement toward the oxygen of trimethylamine oxide than toward the nitrogen of triethylamine. This interpretation is in harmony with the low value reported by Kraus (table 6, footnote e) for the dissociation into ions of the picrate of trimethylamine oxide in nitrobenzene.

A different sample of dehydrated trimethylamine oxide absorbed atmospheric moisture so quickly that the melting point corresponded to the dihydrate. Consequently, it is not certain that the sample that was mixed with bromphthalein magenta in benzene was free of water. No visual change was evident in the solution after several months. A striking change occurred when methyl or ethyl alcohol was added. One drop of the alcohol caused a marked fading of the magenta color, and several drops destroyed it completely, leaving the yellow color of the indicator acid. *n*-Butyl alcohol and cyclohexanol produced a smaller effect, and *iso*propyl alcohol was still less effective. Water was without visual effect, as were acetone, ethyl acetate, and dioxane. Pending further experimental work, it cannot be said whether the effect was due to the alcohol itself or to a trace of an acidic impurity. This behavior of the salt of trimethylamine oxide with bromphthalein magenta makes it readily distinguishable from solutions that contain the salt of bromphthalein magenta with unoxidized tertiary amines, such as trimethyl- or triethylamine.

10. Reaction of Bromphthalein Magenta with Organic Bases in Other Organic Solvents. Discussion of Solvent Effects

Cyclohexane.—In this solvent, as in benzene, bromphthalein magenta gives different colors with primary, secondary, and tertiary amines. However, the salts formed have limited solubilities, as shown by the fact that a Tyndall beam could be observed when more than a minute quantity of the amine was present. Transmittancy curves for bromphthalein magenta B, 5×10^{-5} *M* in cyclohexane (curve 1), and for its products with diethylamine (curve 8) and with triethylamine (curves 2 through 7) are shown in figure 23. A

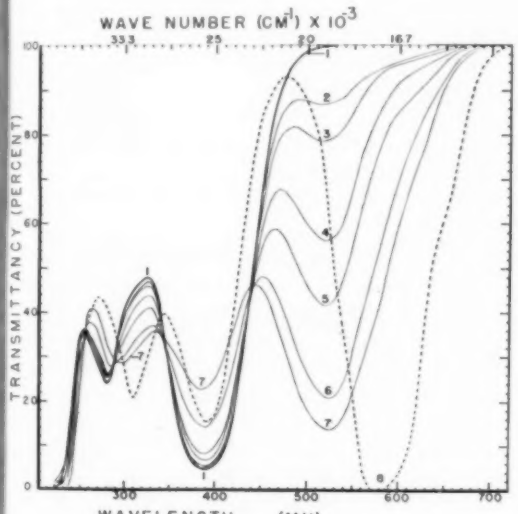


FIGURE 23. (1 to 7) *Transmittancy curves for bromphthalein magenta B, 5×10^{-5} M in cyclohexane, in 1-cm cell, with various concentrations of triethylamine expressed as molar equivalents of the indicator.*

(1) None; (2) 1; (3) 2; (4) 5; (5) 8; (6) 20; (7) 40; (8) transmittancy curve for same solution of the indicator with 1 molar equivalent of diethylamine, 1-cm cell.

1-cm absorption cell was used throughout. Curve 8 was obtained for the indicator after the addition of one molar equivalent of diethylamine. Curves 2 through 7 were obtained for solutions that contained the following successive concentrations of triethylamine, expressed as molar equivalents of the indicator acid: (2) 1; (3) 2; (4) 5; (5) 8; (6) 20; (7) 40. A Tyndall beam was not detected for the solutions that gave curves 2, 3, and 4. With 20 molar equivalents of triethylamine, a Tyndall beam appeared after the solution had stood overnight, and with 40 molar equivalents, immediately after preparation of the solution. Maximum absorbancy in the visible region occurs 15 to 20 mμ nearer the ultraviolet in cyclohexane than in benzene. The transmittancy curves show spreading at the higher wavelengths, probably due to scattering of light by suspended particles.

Ethylene dichloride.—In figure 24 are shown transmittancy curves for bromphthalein magenta E, 5×10^{-5} *M* in ethylene dichloride (curve 1), and for its reaction products with one molar equivalent of the following amines: (2) *n*-Butylamine; (3) di-*n*-butylamine; (4) triethylamine; (5) 1,3-di-*o*-tolylguanidine. Curve 6 was obtained for the 5×10^{-5}-*M* solution of the tetraethylammonium salt of bromphthalein magenta *E* in ethylene dichloride. A 5-mm absorption cell was used throughout.

The behavior of the tetraethylammonium salt in different solvents, including ethylene dichloride, was discussed above (sections IV–5 and IV–8), and it was concluded that the color of the anion shows no significant interaction of the anion with the cation and no appreciable solvent effect. The situation is different for the incompletely alkylated ammonium salts. As in benzene, these salts behave differently from the tetraalkylammonium salts. However, in ethylene dichloride primary, secondary, and tertiary ammonium salts show similar behavior instead of the characteristic differences observed for solutions in benzene. From figure 24 it is evident that in ethylene dichloride the *n*-butylammonium, di-*n*-butylammonium, triethyl-ammonium, and 1,3-di-*o*-tolylguanidinium salts of bromphthalein magenta possess similar transmittancy curves, with a maximum absorbancy (minimum transmittancy) near 570 mμ. This "leveling action" of ethylene dichloride may be explained as analogous to the effect of oxygen compounds (see fig. 19, II). In other words, a chlorine atom in ethylene dichloride, like oxygen in alcohols, ethers, etc., apparently can serve as a proton acceptor with primary or secondary ammonium salts, with the formation of a hydrogen bridge between N and Cl. As a result, the dimeric form of the salt gives place to the monomer. An additional effect of ethylene dichloride is to shift the absorption band approximately 30 mμ toward the infrared. A bathochromic effect upon dyes is often observed as the dielectric

755528—47——5

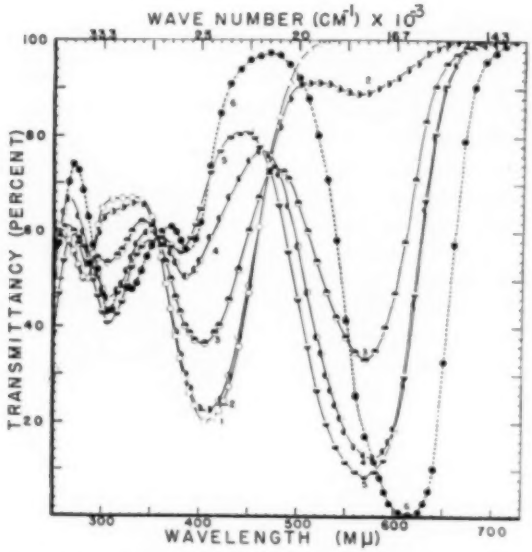


FIGURE 24. (1 to 5) *Transmittancy curves for bromphthalein magenta E, 5×10^{-5} M in ethylene dichloride, in 5-mm cell, with 1 molar equivalent of the following bases.*

(1) None; (2) n-butylamine; (3) di-n-butylamine; (4) triethylamine; (5) 1,3-di-o-tolylguanidine; (6) transmittancy curve for tetraethylammonium salt of bromphthalein magenta E, 5×10^{-4} *M* in ethylene dichloride, 5-mm cell.

constant of the solvent is increased, and the dielectric constant of ethylene dichloride (10.24) is appreciably greater than that of benzene (2.27). Whether ethylene dichloride exerts its effect through dipole orientation [76] or in some other manner is debatable. The apparent effect of the solvent, however, is to reduce the extent of interaction of the indicator anion with the various cations, thus increasing the degeneracy of the resonance structures of the anion.

The reactions shown by curves 2 through 5 were incomplete, as only one molar equivalent of amine had been added to the indicator. The transmittancies at 570 mμ are therefore a measure of the relative strengths of the bases. It is evident that these basicities fall in the same order as in benzene solution.

Other solvents.—Discussion of solvent effects.— Qualitative observations of the color of bromphthalein magenta and of its salts with primary, secondary, and tertiary amines in solvents of different types are summarized in table 7. For the indicator and for these three types of salts, a solvent effect is plainly evident. In the table the color of the quaternary ammonium salt is also stated for a number of solvents; for this type of compound a solvent effect is absent or slight. Conclusions drawn from visual tests are necessarily tentative because spectrophotometric measurements often reveal differences that are not evident to the eye. It is also possible that traces of impurities were present in some of the solvents. Furthermore, as stated in section IV–8, in oxygen-containing solvents, the color is dependent upon the concentration of the indicator and base. Nevertheless, for bromphthalein magenta and for its quaternary or tertiary ammonium salts, the following simple explanations of solvent effects are possible:

TABLE 7.—*Effect of solvent and added base upon the color of bromphthalein magenta* [a]

Solvent [b]	Type of base added [c]				
	0	I	II	III	IV
A. APROTIC SOLVENTS					
Benzene (2.27)	Yellow	Red-purple	Blue [d]	Magenta	Blue.[e]
Carbon tetrachloride (2.22)	do	do	do [d]	do	Do.[e]
Cyclohexane (2.01)	do	do	do [d]	do	Do.[e]
Toluene (2.31, 18 ° C)	do	do	do [d]	do	Do.[e]
Carbon disulfide (2.63, 20° C)	do	Magenta [f]	Magenta [f]	do	Do.[e]
B. PARTLY HALOGENATED HYDROCARBONS					
Chlorobenzene (5.61)	Yellow	Red-purple	Red-purple	Magenta	Blue.[e]
Chloroform (4.72)	do	do	do	do	Do.[e]
Ethylene dichloride (10.24)	do	Magenta	Magenta	do	Do.[e]
C. ETHERS					
Dioxane (2.23)	Yellow	Magenta	Magenta	Magenta	Blue.[e]
Diethyl ether (4.26)	do	do	do	do	Do.[e]
Diethyl "Cellosolve"	do	Red-purple	Red-purple	do	
Anisole (4.29)	do	Magenta	Magenta	do	Blue.[e]
D. ESTERS					
Ethyl acetate (4.5, 18° C)	Green-yellow	Red-purple	Red-purple	Magenta	Blue.[e]
Butyl stearate (3.12, 30° C)[g]	Yellow	Magenta	Magenta	do	
Dibutyl sebacate (4.55, 30° C)[g]	do	do	do	do	
Methyl benzoate (6.9, 20° C)	Green-blue	Red-purple	Red-purple	do	
Benzyl benzoate (4.8, 20° C)	Yellow	do	do	do	
Tributyl phosphate (7.97, 30° C)[g]	Blue-green	Blue [e]	Blue [e]	Blue [e]	
E. KETONES					
Acetone (20.88)	Yellow to blue [h]	Blue [e]	Blue [e]	Blue [e]	Blue.[e]
Methyl isobutyl ketone	do [h]	do [e]	do [e]	do [e]	
Acetonylacetone	do [h]	do [e]	do [e]	do [e]	
Cyclohexanone (18.2, 20° C)	Yellow to green [h]	do [e]	do [e]	do [e]	
F. ALCOHOLS [i]					
Ethyl alcohol (23.2, 30° C)	Blue [e]	Blue [e]	Blue [e]	Blue [e]	Blue.[e]
n-Propyl alcohol (21.2, 21.9° C)	Green-blue	do [e]	do [e]	do [e]	
n-Butyl alcohol (18.4, 20.5° C)	Green	do [e]	do [e]	do [e]	
Heptanol-2	Green-yellow	do [e]	do [e]	Red-purple	
Octyl alcohol (2-ethylhexanol)	Yellow	Red-purple	Red-purple	do	
Heptadecanol (3,9-diethyltridecanol-6)	Magenta	do	do	Magenta	
Cyclohexanol (8.5, 18° C)	Blue-green	Blue [e]	Blue [e]	Blue [e]	
Benzyl alcohol (13.6, 21.5° C)	Green	do [e]	do [e]	do [e]	
2-Methyl-2,3-pentanediol (24.4, 30° C) [g]	Blue [e]	do [e]	do [e]	do [e]	

See footnotes at end of table.

TABLE 7.—*Effect of solvent and added base upon the color of bromphthalein magenta* [a]—Continued

Solvent [b]	Type of base added [c]				
	0	I	II	III	IV
G. MISCELLANEOUS					
Acetonitrile (36.7)	Green-blue	Blue [e]	Blue [e]	Blue [e]	
Aniline (7.3, 21.6° C)	do	do [d]	do [d]	do [d]	Blue [e]
Nitrobenzene (34.9)	Yellow	Green	Green-blue	Green	Do [e]
Nitromethane (35.76, 30° C) [g]	do	Blue [d]	Blue [e]	Blue [e]	
Pyridine (12.5, 20° C)	Green-blue	Green-blue	do [e]	do [e]	

[a] Bromphthalein magenta B was used in these tests.
[b] Numbers in parentheses are dielectric constants, at 25° C unless otherwise indicated. Values were taken from the literature [92,93], except as noted.
[c] Key to bases used: O=no base; I=$C_8H_{17}NH_2$; II=$(n\text{-}C_4H_9)_2NH$; III=$(C_2H_5)_3N$, IV=$(n\text{-}C_4H_9)_4N$—salt of bromphthalein magenta B.
[d] Purple-blue.
[e] Blue, without purple tinge.
[f] Base reacted to some extent with solvent.
[g] Values were obtained by A. A. Maryott and C. G. Malmberg of the Bureau for high-grade commercial solvent.
[h] Yellow in concentrated solution, bluer upon dilution.
[i] In concentrated solutions, the color shifts toward yellow.

(1) *Bromphthalein magenta.*—As stated in the introduction, it is fairly well established that ionization of most acids is dependent chiefly upon the presence of proton-attracting molecules. For example, Kraus observed that hydrogen chloride is a very weak electrolyte in nitrobenzene, in spite of the rather high dielectric constant of this solvent [57a]. On the other hand, perchloric acid is a rather strong electrolyte in nitrobenzene. There is evidently a considerable difference in the affinity of the chloride and perchlorate ions for protons. As pointed out in the introduction, this difference is masked in aqueous solutions, as protons of both acids unite with the solvent to give H_3O^+ ions. From the behavior of the acid, bromphthalein magenta, it is to be inferred that the affinity of its anion for a proton is high. In solvents of classes A, B, and C, there is no evidence of ionization, whereas solvents that have a high affinity for protons, such as alcohols and pyridine, compete with the anion of the acid, producing solvated hydrogen ions and liberating the blue anions. Ionization of the indicator acid is incomplete, however, except at very high dilutions. The peculiar behavior shown by heptadecanol (3,9-diethyltridecanol-6) in producing a magenta tone rather than a blue color was also observed for tetradecanol (7-ethyl-2-methylundecanol-4). It is noteworthy that both compounds have two large hydrocarbon radicals adjacent to the —OH group. The following explanation for the magenta color is suggested: The hydrogen in the —OH group of the indicator acid forms a bridge to the oxygen in the —OH group of the alcohol, with the alcohol taking the role of a base. The hydrogen bridge is shielded by the hydrocarbon radicals in the alcohol, so that further solvation of the proton is impeded. It seems likely that a completely solvated proton forms a hydrogen bridge between two molecules of solvent, and complete solvation by either of these alcohols may be prevented by steric factors.

(2) *Quaternary ammonium salts.*—For these compounds, the attraction between the cation and the anion is Coulombic in nature and probably has little or no influence upon the electronic system of the anion. On the basis of the resonance hypothesis, the degeneracy of the structures contributing to the resonance of the anion may be considered essentially complete. Consequently, the extent to which a quaternary ammonium salt exists as free ions, ion-pairs, or other aggregates is not measurable spectrophotometrically.

(3) *Tertiary ammonium salts.*—In section IV-4, the behavior of tertiary ammonium salts in benzene was considered and a hydrogen bridge between the cation and anion was postulated. This is the explanation that Kraus advanced to account for the low dissociation constants of most salts of this type in solvents having no affinity for the proton. He attributed the leveling effect of pyridine upon the dissociation of ammonium salts to interaction of the proton of the cation with a pyridine molecule. In the absence of a proton-

attracting solvent, the proton of the cation interacts strongly with the anion of the salt [57]. The behavior of trialkylammonium salts of bromphthalein magenta may be given a similar interpretation. That is, a magenta color indicates interaction between the proton of the trialkylammonium ion and the anion; a blue color is evidence that the proton of the cation has interacted with a molecule of the solvent. If the conductance were measured, a negligible value for the magenta solution and a measurable value for the blue solution would be predicted. In analogy with conclusions reached by Brooker in studies of other dyes [77], one may conclude that the degeneracy of the resonance structures of the anion in a trialkylammonium salt of bromphthalein magenta is incomplete, as shown by the fact that spectral absorption occurs at shorter wavelengths than that of tetraalkylammonium salts. It is to be recalled that salts of symmetrical diarylguanidines and triphenylguanidine behave in the same way as trialkylammonium salts of bromphthalein magenta.

(4) *Secondary ammonium salts.*—Before discussion of dialkylammonium salts of bromphthalein magenta, the experimental data given in table 7 and in earlier sections of the paper may be summarized:

(a) In hydrocarbons and in carbon tetrachloride, the color is purple-blue; (b) addition of small amounts of alcohol or other oxygen-containing solvents produces a change in color from purple-blue to magenta; (c) in partly halogenated hydrocarbons, ethers, and most esters, the color is magenta or red-purple; (d) in solvents of high proton-affinity, the color is blue. This is a different color, visually and spectrophotometrically, from the purple-blue of solutions in solvents of class *A*.

An explanation for observations (a), (b), and (c) was offered in section IV–8, and on page 253. The blue color of solutions in solvents of high proton-affinity (observation (d)) resembles that exhibited by tertiary-ammonium salts in these solvents and is to be accounted for in a similar way.

In addition to the effects discussed in the preceding paragraph, note should be made of the fact that the purple-blue of solutions in benzene is changed to magenta, not only by compounds that contain oxygen or halogen, but also by other types of solvents, including acetonitrile, aniline, and nitrobenzene. In studying the deuteron-attracting properties of liquids by infrared spectroscopy, Gordy and his collaborators [78, 79] measured the shift of the OD band of heavy methyl alcohol, CH_3OD, or heavy water, D_2O. Benzene was used as reference solvent for CH_3OD or D_2O. Nitrobenzene was found to produce a small but measurable shift, and acetonitrile, oxygen compounds, and amines produced a greater shift, in the order named. The shift produced in the OD band was greater for D_2O than for CH_3OD. The OD band in CH_3OD was not shifted by bromobenzene. The conclusion was reached that nitrobenzene can be linked through deuterium to the oxygen of D_2O or CH_3OD. Moreover, an intramolecular hydrogen bridge in *o*-nitrophenol is generally postulated to account for differences in the behavior of this compound and its *m*- and *p*-isomers.

From the experimental data presented in this paper, it is evident that in a dialkylammonium salt of bromphthalein magenta, the cation behaves as a strong proton donor. When the nitrogen of a dialkylamine interacts with the proton of bromphthalein magenta, the proton originally attached to the amine nitrogen evidently becomes more loosely held, so that it interacts readily with oxygen, halogen, or nitrogen atoms of the same or other molecules.[35] As a result, the secondary ammonium salt of bromphthalein magenta becomes either dimerized or solvated, depending upon whether the medium is an inert solvent or contains compounds of high proton-affinity. The salt can be solvated by bromobenzene and chlorobenzene, as well as by the solvents found by Gordy to exhibit bonding with CH_3OD.

The solvent, tri-*n*-butyl phosphate, deserves special comment. From table 7, class D, it is evident that this compound possesses a higher solvating capacity than the usual types of esters. It is therefore interesting to note that it causes a much larger shift in the OD band of CH_3OD than other esters [79]. Moreover, the esters of phosphoric acid are exceptionally good solvents for monofluorodichloromethane [80]. In all these cases, the observations can be attributed to the strong proton-affinity of the P=O group.

[35] As a corollary, it might be postulated that a mono- or dialkylamine should show greater reactivity with an acid (in an aprotic solvent) after the addition of a proton-accepting substance.

From a comparison of curves 2, 3, and 4 in figure 21 it may be concluded that in the dimeric form of a dialkylammonium salt of bromphthalein magenta, the degeneracy of the resonance structures is almost as great as in a tetraalkylammonium salt [77]. Dicyclohexylguanidine appears to resemble secondary amines in its behavior rather than tertiary amines and di- or triarylguanidines.

(5) *Monoalkylammonium salts.*—As stated in section IV–4, monoalkylammonium salts of bromphthalein magenta appear to exist partly in the monomeric and partly in the dimeric condition. The limited stability of the dimeric form may be explained in several ways:

(a) A monoalkylamine does not interact so completely with the proton of the acid and consequently the cation of the salt has weaker proton-donor properties; (b) as the cation possesses two protons, neither proton is so readily donated as the single proton of a dialkylammonium salt; (c) the dimer of a monoalkylammonium salt of bromphthalein magenta has lower symmetry than the dimer of a dialkylammonium salt; (d) a monoalkylamine has sufficiently strong proton-acceptor properties to become coordinated with a mono- or dialkylammonium ion.

Interpretation of the behavior of tetrahydroquinoline involves similar factors, as well as a possible effect of the adjacent benzene ring.

In the above discussion, the possible role of the dielectric constant of the solvent has been to a large extent ignored. As the attraction between the anion and incompletely substituted ammonium ions is partly Coulombic in nature [57], the dielectric constant undoubtedly has some effect. However, the general pattern of behavior of bromphthalein magenta and its salts can be interpreted satisfactorily from a consideration of the proton-affinity of the solvent, but not from a consideration of its dielectric constant. Gordy has pointed out that there is no correlation evident in the dipole moments of solvents and the strengths of the bonds formed with CH_3OD [79].

11. Sulfonephthalein Analogs and Other Indicators

Further brief mention may be made of the fact that the peculiarities in behavior described above for bromphthalein magenta are not characteristic of this indicator alone. For example, differences in the colors produced by the addition of secondary and tertiary amines to the indicator in benzene solution are shown by several other indicators. These include the following sulfonephthaleins: bromcresol green, bromcresol purple, bromphenol blue, chlorphenol blue, bromchlorphenol blue, iodophenol blue, and tetrabromphenol blue. Ethyl eosin (the ethyl ester of eosin free acid) and fluorescent blue show characteristic differences in their behavior with secondary and tertiary amines that are quite marked spectrophotometrically, though less readily detected by the unaided eye [43]. These indicators have a related structure. With the exception of fluorescent blue, they may be regarded as derivatives of benzaurin (fig. 4), with one or more halogen atoms ortho to the phenol and quinoid oxygens.

In inert solvents, bromphthalein magenta is superior to the other indicators listed here, on account of its greater solubility as well as the brilliant and sensitive color changes that it undergoes in reactions with aliphatic amines. Furthermore, being a monobasic acid, its reactions can be studied more readily than those of the sulfonephthalein indicators, which are dibasic acids. As is to be expected, the carbethoxy group, $-COOC_2H_5$, undergoes hydrolysis by strong aqueous acid or alkali and conversion to tetrabromophenolphthalein results. In addition, bromphthalein magenta is almost insoluble in water, as by intention it lacks the water-solubilizing sulfonic acid group that is present in the sulfonephthaleins. In short, an indicator must be tailored to fit the medium in which it is to be used. The sulfonephthaleins are excellent indicators for aqueous media, but unsuited for use in hydrocarbons, whereas the reverse is true for bromphthalein magenta.

Bromphthalein magenta, the analog of bromphenol blue, is only one of a series of indicators that should prove useful in organic media. The analogs of bromcresol green and other sulfonephthaleins would undoubtedly be valuable. Other derivatives of benzaurin, not necessarily containing the carbethoxy group, are possibilities. For example, α-naphtholbenzein is soluble in benzene and, being less acidic than bromphthalein magenta, is more satisfactory for some purposes.

The progressive solvation of incompletely alkylated ammonium ions, so readily observed for the salts of bromphthalein magenta and the other halogenated indicators referred to above, because of the striking color changes that accompany

ation, undoubtedly occurs in the case of alkylammonium salts of other acids. This study of bromphthalein magenta therefore serves not only to clarify its behavior and that of its analogs, but to aid in interpreting the behavior of alkylammonium salts in general.

12. Acid-base Titrations in Organic Solvents

For many years, it was usually assumed that acid-base reactions could not occur in a non-aqueous medium. Hantzsch and numerous others showed that this supposition is incorrect. Moreover, the feasibility of performing quantitative colorimetric titrations in inert solvents has been reported by several investigators. The progress made in this field may be reviewed briefly.

Aminometric titrations.—In 1903, Vorländer announced the successful titration of aniline with a dilute solution of anhydrous hydrogen chloride in benzene, with methyl yellow (*p*-dimethylaminoazobenzene) as indicator [81]. At that time, Vorländer was interested in correlation of the color and structural changes of the indicator rather than in titrimetric analyses. Thirty years passed before he turned to possible analytical applications of the titration, for which he coined the name "aminometry" or "alkalimetry without hydroxyl ions" [82, 83]. Investigations showed that with chloroform or carbon tetrachloride in place of benzene as the solvent, the color change was sharper. Chloroform was best because of its solvent power and could be used alone or in mixtures with benzene or toluene. Benzenesulfonic acid and other anhydrous sulfonic acids were as satisfactory as hydrogen chloride, and changed color less rapidly. The acid could be standardized against aqueous alkali or against tribenzylamine in chloroform.

A number of aromatic and aliphatic amines not satisfactorily titratable in aqueous or alcoholic media could be titrated with results ranging from less than 1 percent to 2 percent below the theoretical values. Volatile amines were dissolved in an excess of the acid and back-titrated with tribenzylamine solution. Brucine, one of the alkaloids that titrated successfully, gave equally good results for the anhydrous and the hydrated forms. With precautions, both amino groups of certain diacidic alkaloids could be titrated. Picric acid was not usable for lack of a suitable indicator.

A short time later, Dietzel and Paul became interested in Vorländer's aminometric procedure as a means of simplifying pharmacopoeial methods for the analysis of alkaloids and other drugs and their preparations and made further tests of its applicability to primary, secondary, and tertiary aliphatic and aromatic amines [84]. In preliminary experiments, the titration vessel was protected from access of atmospheric moisture, but this precaution was later found unnecessary. Anhydrous *p*-toluenesulfonic acid in chloroform proved to be the most satisfactory standard solution. Hexamethylenetetramine, which acts as a monoacid base, was equally as good as tribenzylamine for determining the titer of the acid, was cheaper and was more easily purified. The base concerned was weighed exactly into a beaker and titrated with 0.05-*M* *p*-toluenesulfonic acid in chloroform, with the addition of 5 drops of 0.05-percent methyl yellow in chloroform until the color changed sharply to red. Alkaloids and drugs that could be titrated with an accuracy of ± 1 percent included atropine, cinchonine, papaverine, narcotine, codeine, strychnine, and ephedrine. The addition of about 1 percent of chemically pure phenol by volume led to a sharper color change. Morphine, although insoluble in chloroform alone, dissolved in 1:3 phenol-chloroform and was titrated exactly. Accurate results were obtained for nonvolatile aliphatic amines and for aromatic amines of basicity comparable to that of aniline. Any halogen-containing solvent, such as tetrachloroethane or methylene chloride, could be substituted for chloroform.

Khalt [85] used the procedure of Dietzel and Paul in the assay of atropine, lobeline, and harmine. Preliminary experiments with various types of amines confirmed the superiority of *p*-toluenesulfonic acid to other acids. A 0.05-*M* solution of the acid in chloroform gave better results than a less concentrated solution. With bromphenol blue as the indicator, the color change was not sharp in many cases. Quinine was found to titrate as a monoacidic base in the presence of bromphenol blue and as a diacidic base in the presence of methyl yellow. Tribenzylamine was not available for standardization of the acid, and codeine recrystallized from alcohol and dried to constant weight at 100° C proved a satisfactory substitute. The results were much better when chloroform was used as the solvent than when the

titrations were performed in aqueous or alcoholic media.

Trautner and coworkers have made further applications of Vorländer's aminometric titration to the analysis of pharmaceutical preparations and plant materials for their alkaloid content, following the modifications introduced by Dietzel and Paul [86, 87]. *p*-Toluenesulfonic acid suitable for the analyses is obtained by dehydrating the purified monohydrate. This is accomplished by heating the monohydrate at 10-mm pressure and under a current of air dried with phosphorus pentoxide for 1 hour to 120° C. The product is a grayish crystalline mass that melts at 105° C, is not very hygroscopic, and can be kept in a well-stoppered bottle [86]. Nicotine, which is slightly more than monoacidic toward *p*-toluenesulfonic acid, reacts as a diacidic base with picric acid. Precipitation of the monopicrate is prevented by warming the solution to about 40° C [87].

In nearly all of the experiments cited, the indicator used was methyl yellow. Methyl yellow is probably a stronger base than some of the bases titrated, and this may account for the fact that in a number of cases less acid was consumed than that equivalent to the known content of base. The weaker base, benzeneazodiphenylamine, dissolves readily in inert solvents and changes color upon the addition of sulfonic acids. Its use in place of methyl yellow might yield more accurate results in titrations of some of the weaker organic bases. There are probably still other indicator dyes, commercially available or readily synthesized, which would extend the usefulness of the method. The role of phenol in sharpening the methyl yellow endpoint is not explained and deserves investigation. *p*-Toluenesulfonic acid appears to be the most satisfactory reference acid for most purposes, but at the present time it is necessary for each worker to purify and dehydrate the commercial product. A wider recognition of the feasibility of such titrations and commercial availability of anhydrous strong acids and suitable indicators should lead to other analytical applications.

Titrations of weak organic acids in mixtures of absolute alcohols with aprotic solvents.—About 35 years ago, Folin and coworkers directed attention to the successful titration of weak acids, such as benzoic and stearic acids, dissolved in benzene, toluene, chloroform, or carbon tetrachloride, with sodium ethylate or amylate in the corresponding absolute alcohol as the base, and any one of several indicators, preferably phenolphthalein. Although the conductivities of their solutions were found to be extremely low, the reactions were instantaneous and the behavior of the indicator essentially the same as in aqueous solutions [88 to 90]. This work did not receive wide attention, as it antedated the Brönsted-Lowry and Lewis theories and the results reported violated the theory then current, that dissociation into ions is a prerequisite in chemical reactions.

The Folin method is excellent for determination of the acid content of impure solvents, such as ligroin and turpentines, except in the case of very old turpentines in which extremely weak acids, possibly peroxides, show a buffering effect upon the endpoint [91]. This might be overcome by the use of an indicator of higher acidity than phenolphthalein. Bromphthalein magenta can be used in the titration of relatively strong acids, such as trichloroacetic acid, but is too strong an acid to be used in the titration of acetic and benzoic acids. Dicyclohexylguanidine, 0.1 *M* in absolute ethyl alcohol, is a convenient substitute for sodium ethylate and gives exact results in the titration of trichloroacetic acid, picric acid, or benzoic acid in benzene, with α-naphtholbenzein as indicator but not with phenolphthalein. Bromphthalein magenta is slightly superior to α-naphtholbenzein in the titration of picric acid [91].

The Folin procedure was adapted by Lavine and Toennies to the differential titration of acetic and perchloric acids in acetonitrile, with methyl alcoholic sodium methylate as the base. Thymol blue was used as the indicator, changing from pink to yellow when the perchloric acid was neutralized, and from yellow to blue at the acetic acid endpoint [92].

Titrations in aprotic solvents with bromphthalein magenta as indicator.—As shown earlier in this paper, 1,3-diphenyl-, di-*o*-tolyl- and dicyclohexylguanidine appear to be stronger bases in benzene than any of the aliphatic amines, with increasing strength in the order named. Dicyclohexylguanidine is too difficultly soluble in benzene for practical use, although it produces a color change not only with bromphthalein magenta but with the weaker acid, α-naphtholbenzein. Di-*o*-tolyl- and diphenylguanidine react readily with bromphthalein magenta but not with α-naphtholbenzein. As

bromphthalein magenta is a relatively strong acid, it cannot be used in the titration of acids as weak as benzoic and acetic acids. As pointed out earlier in the discussion, there is need for a strong guanidine derivative or other organic base that is readily soluble in benzene. Dicyclohexylguanidine is fairly soluble in chloroform but reacts with the solvent, particularly upon warming. This reaction is not sufficiently rapid at room temperature to prevent excellent titrations with acetic acid, *o*-chlorophenoxyacetic acid, and picric acid, with *α*-naphtholbenzein as the indicator [91].

A series of experimental titrations with benzene as solvent and a 0.1-percent solution of bromphthalein magenta in benzene as the indicator is summarized in table 8. The total volume was approximately 30 ml, and the volume of acid or base was less than 10 ml in all cases and as little as 1 to 2 ml in some instances. Moreover, in three of the titrations the concentrations of the acid and the base differed by a factor of two. The endpoints were very sharp, vivid, and reproducible except in the titration of piperidine with trichloroacetic acid, where the formation of a cloud caused slightly lower precision.

From inspection of the table it can be noted that trichloroacetic acid and picric acid, used at different concentrations, gave the same value in the titration of ditolylguanidine. With triethylamine or piperidine as the base, trichloroacetic acid gave lower values than picric acid. From the standpoint of reproducibility and vividness of the color change, the results are very satisfactory.

Diphenylguanidine was used by Kilpatrick for the quantitative estimation of the very strong acid, perchloric acid, in acetonitrile, with methyl yellow, benzoyl auramine, or thymol blue as the indicator, but an organic base sufficiently strong for the estimation of acetic acid in this solvent was lacking [45]. It should be possible, after further investigations, to obtain strong acids and bases and a series of indicators that will permit the quantitative titration of weak bases and weak acids in the various aprotic solvents.

TABLE 8.—*Acid-base titrations in benzene with bromphthalein magenta as indicator*

Reference base	Acid	Molarity found for acid
Diphenylguanidine, 0.0100 *M*.[a]	*d*-Camphorsulfonic acid (Reychler), approximately 0.005 M.	0.00527, 0.00529, 0.00528.
Reference acid	**Base**	**Molarity found for base**
d-Camphorsulfonic acid, 0.00528 *M*.[b]	Impure dicyclohexylguanidine, approximately 0.0025 *M*.	0.00238, 0.00237
Trichloroacetic acid, 0.106 *M*.[c]	Di-*o*-tolylguanidine, approximately 0.05 *M*.	0.0484, 0.0484, 0.0487.
Picric acid 0.0100 *M*[c]	Di-*o*-tolylguanidine, approximately 0.01 *M*.[d]	0.00970, 0.00972, 0.00972.
Trichloroacetic acid, 0.106 *M*.[c]	Triethylamine, approximately 0.1 *M*.[e]	[f] 0.0942, 0.0944.
Trichloroacetic acid, 0.0106 *M*.[g]	Triethylamine, approximately 0.01 *M*.[h]	0.00935, 0.00935, 0.00936.
Picric acid, 0.0100 *M*[e]	....do[h]	0.00962, 0.00962, 0.00960, 0.00962.
Trichloroacetic acid, 0.106 *M*.[c]	Piperidine, approximately 0.1 *M*.[e]	[f] 0.0957, 0.0953.
Picric acid, 0.100 *M*[e]	....do[e]	[i] 0.0982, 0.0990, 0.0977.

[a] The solid was assumed to be 100 percent pure.
[b] Standardized against diphenylguanidine in benzene.
[c] Standardized against sodium ethylate in absolute ethyl alcohol with phenolphthalein as indicator; the sodium ethylate was standardized against benzoic acid with the same indicator.
[d] Diluted from the approximately 0.05-*M* solution.
[e] Would be 0.100 *M* if no impurities were present.
[f] A cloud formed during the titration. Methyl yellow did not give a sharp color change.
[g] Diluted from the 0.106-*M* solution.
[h] Diluted from the approximately 0.1-*M* solution.
[i] Picrate began to precipitate out.

V. Summary. Applications of Bromphthalein Magenta

The differentiating reaction of bromphthalein magenta with various classes of organic bases in inert solvents, and its uses for quantitative comparisons of basic strengths and for certain acid-base titrations were described above. An additional application is in the measurement of the relative strengths of organic acids in inert solvents. Studies are in progress in this laboratory in which the strength of an acid is measured by the extent of its competition with bromphthalein magenta for a reference base, such as di-*o*-tolylguanidine. Preliminary measurements have been made for more than 30 acids, that include acetic and benzoic acids as well as a number of their more highly acidic derivatives. Some of these are too insoluble to be studied in aqueous solutions [91].

Other applications of the indicator to specific analytical problems will doubtless occur to some readers. The description and interpretations given in this paper of the behavior of the indicator acid in different environments will serve as a guide in selecting the proper conditions. As indicated

in section IV–9, qualitative tests can be made with salts as well as with the corresponding bases.

The experimental data of the authors and the investigations of other workers that have been reviewed in this report point to promising lines of work in the practical and theoretical study of acid-base reactions.

The authors express their indebtedness to S. F. Acree, former chief of section V–10, pH Standards, for suggesting the development of new indicators for the measurement of acidity and basicity in organic solvents, for his interest in the progress of the work, and for his assistance in the designing of various absorption cells used in the spectrophotometric measurements. Grateful acknowledgment is also made to Kenneth D. Fleischer for the chemical analyses, to Wilbur W. Brannon and Fred H. Engler for construction of the stoppered endplates and other portions of the absorption cells, to Mary Ellen Lovelace for assistance in part of the synthetic work, and to the commercial organizations listed in section III of this report for their contributions of bases and solvents used in the investigations.

VI. References

[1] J. N. Brönsted, Rec. trav. chim. **42,** 718 (1923).
[2] J. N. Brönsted, Chem. Rev. **5,** 231 (1928).
[3] J. N. Brönsted, Z. angew. Chem. **43,** 229 (1930).
[4] J. N. Brönsted, Z. physik. Chem. [A] **169,** 52 (1934).
[5] T. M. Lowry, Trans. Faraday Soc. **20,** 13 (1924).
[6] A. Hantzsch, Z. Elektrochem. **29,** 221 (1923); **30,** 194 (1924).
[7] A Hantzsch and W. Voigt, Ber. deut. chem. Ges. **62,** 975 (1929).
[8] N. F. Hall and J. B. Conant, J. Am. Chem. Soc. **49,** 3047 (1927).
[9] N. F. Hall and T. H. Werner, J. Am. Chem. Soc. **50,** 2367 (1928).
[10] N. F. Hall, J. Am. Chem. Soc. **52,** 5115 (1930).
[11] E. C. Franklin, The nitrogen system of compounds (Reinhold Publishing Corp., New York, N. Y., 1935).
[12] T. S. Moore, J. Chem. Soc. (London) **91,** 1373, 1379 (1907).
[13] W. M. Latimer and W. H. Rodebush, J. Am. Chem. Soc. **42,** 1419 (1920).
[14] N. V. Sidgwick, The electronic theory of valency (Oxford Univ. Press, New York, N. Y., 1927).
[15] M. L. Huggins, J. Org. Chem. **1,** 407 (1936).
[16] E. N. Lassettre, Chem. Rev. **20,** 258 (1937).
[17] L. Pauling, The nature of the chemical bond (Cornell Univ. Press, Ithaca, N. Y., 1st ed., 1939; 2nd ed., 1940).
[18] W. H. Rodebush and A. M. Buswell, J. Phys. Chem. **43,** 219 (1939); and other papers by Rodebush, Buswell, and collaborators.
[19] H. Hoyer, Z. Elektrochem. **49,** 97 (1943).
[20] W. Gordy, Sci. Monthly **53,** 440 (1941); and numerous papers by Gordy and coworkers in J. Chem. Physics and J. Am. Chem. Soc.
[21] G. Briegleb, Z. Elektrochem. **50,** 35 (1944).
[22] L. Hunter, Chemistry & Industry, p. 155 (1944).
[23] G. W. Wheland, The theory of resonance and its application to organic chemistry (John Wiley & Sons, Inc., New York, N. Y., 1944).
[24] G. N. Lewis, Valence and the structure of atoms and molecules (Chemical Catalog Co., New York, N. Y., 1923).
[25] G. N. Lewis, J. Franklin Inst. **226,** 293 (1938).
[26] G. N. Lewis and G. T. Seaborg, J. Am. Chem. Soc. **61,** 1886 (1939).
[27] W. F. Luder and S. Zuffanti, The electronic theory of acids and bases (John Wiley & Sons., Inc., New York, N. Y., 1946). Much of this material is contained in a series of papers by Luder and coauthors in J. Am. Chem. Soc., J. Chem. Education, and Chem. Rev.
[28] P. Walden, Salts, acids, and bases: Electrolytes: Stereochemistry (McGraw-Hill Book Co., Inc., New York, N. Y., 1929).
[29] Acids and bases (J. Chem. Education, Easton, Pa., 1941); More acids and bases (same publisher, 1944). Both are collections of papers by N. F. Hall, L. P. Hammett, W. F. Luder and others, which originally appeared in J. Chem. Education.
[30] I. M. Kolthoff, J. Phys. Chem. **48,** 51 (1944).
[31] H. H. Cary and A. O. Beckman, J. Optical Soc. Am. **31,** 682 (1941).
[32] A. Baeyer, Liebigs Ann. Chem. **201,** 80 (1880).
[33] R. Nietzki and E. Burckhardt, Ber. deut. chem. Ges. **30,** 175 (1897).
[34] M. E. Maclean, P. J. Jencks, and S. F. Acree, J. Research NBS **34,** 271 (1945) RP1643.
[35] E. Eigenberger, J. prakt. Chem. **130,** 75 (1931).
[36] A. Thiel and R. Diehl, Sitzungsber. Ges. Beförderung gesamt. Naturwissenschaften Marburg **62,** 472 (1927).
[37] P. Ramart-Lucas, Bull. soc. chim. **8,** 865 (1941).
[38] P. Ramart-Lucas, Bull. soc. chim. **10,** 279 (1943).
[39] F. Feigl and V. Anger, Mikrochem. Acta **2,** 107 (1937); M. Ishidate and T. Sakaguchi, Ber. deut. chem. Ges. **74,** 163 (1941).
[40] J. N. Brönsted, Ber. deut. chem. Ges. **61,** 2049 (1928).
[41] V. K. LaMer and H. C. Downes, J. Am. Chem. Soc. **55,** 1840 (1933).
[42] V. K. LaMer and H. C. Downes, Chem. Rev. **13,** 47 (1933).
[43] Marion Maclean Davis, Priscilla J. Schuhmann, and Mary Ellen Lovelace, National Bureau of Standards. Unpublished data.
[44] D. C. Griffiths, J. Chem. Soc. (London), 818 (1938).
[45] M. Kilpatrick, Jr. and M. L. Kilpatrick, Chem. Rev. **13,** 131 (1933).
[46] A. Baeyer, Liebigs Ann. Chem. **354,** 152 (1907).

[47] S. F. Acree and E. A. Slagle, Am. Chem. J. **42,** 115 (1909).
[48] E. Q. Adams and L. Rosenstein, J. Am. Chem. Soc. **36,** 1452 (1914).
[49] C. R. Bury, J. Am. Chem. Soc. **57,** 2115 (1935).
[50] G. Schwarzenbach, M. Brandenberger, G. H. Ott, and O. Hagger, Helv. chim. Acta **20,** 490 (1937).
[51] B. Cohen, Reprint No. 1131 from the Public Health Reports (Government Printing Office, Washington, D. C., 1927).
[52] L. Pauling, J. Am. Chem. Soc. **58,** 94 (1936).
[53] O. R. Wulf, U. Liddel, and S. B. Hendricks, J. Am. Chem. Soc. **58,** 2287 (1936).
[54] J. L. Hague, Proc. Am. Soc. Testing Materials **44,** 712 (1944).
[55] G. Briegleb, Naturwiss. **30,** 436, 469, 506 (1942).
[56] T. S. Moore and T. F. Winmill, J. Chem. Soc. (London) **101,** 1635 (1912).
[57a] C. A. Kraus, J. Phys. Chem. **43,** 231 (1939).
[57b] C. A. Kraus, Science **90,** 281 (1939), and numerous other publications by Kraus and coworkers.
[58] N. L. Cox, C. A. Kraus, and R. M. Fuoss, Trans. Faraday Soc. **31,** 749 (1935).
[59] E. C. Lingafelter, J. Am. Chem. Soc. **63,** 1999 (1941).
[60] H. C. Brown, H. I. Schlesinger, and S. Z. Cardon, J. Am. Chem. Soc. **64,** 325 (1942); H. C. Brown and H. Bartholomay, Jr., J. Chem. Phys. **11,** 43 (1943); H. C. Brown, Science **103,** 385 (1946); and other papers by H. C. Brown and coworkers.
[61] A. Thiel, A. Dassler, and F. Wülfken, Fortschr. Chem. Physik physik. Chem. **18,** (3), 1 (1924).
[62] W. M. Clark, The determination of hydrogens ions, p. 161 (Williams and Wilkins Co., Baltimore, Maryland, 3rd ed., 1928).
[63] I. M. Kolthoff, Acid-base indicators, p. 122 (translation from the 4th German ed. by C. Rosenblum, Macmillan Co., New York, N. Y., 1937).
[64] E. Heuser, Organic chemistry, an advanced treatise, Henry Gilman, Editor-in-Chief, **II,** 1672 (John Wiley & Sons, Inc., New York, N. Y., 2nd ed., 1943).
[65] R. M. Fuoss and C. A. Kraus, J. Am. Chem. Soc. **55,** 3614 (1933); discussed by H. S. Harned and B. B. Owen, The physical chemistry of electrolytic solutions (American Chemical Society Monograph No. 95, Reinhold Publishing Corp., New York, N. Y., 1943).
[66] W. H. Zachariasen, J. Am. Chem. Soc. **54,** 3841 (1932).
[67] B. E. Warren, Chem. Rev. **26,** 237 (1940).
[68] L. Hunter and J. A. Marriott, J. Chem. Soc. (London), 777 (1941).
[69] N. Bjerrum, Kgl. Danske Videnskab. Selskab, Math.-fys. Medd. **7,** 3 (1926).
[70] W. Gordy, J. Am. Chem. Soc. **59,** 464 (1937).
[71] A. A. Maryott, National Bureau of Standards. Unpublished data.
[72] T. L. Davis and R. C. Elderfield, J. Am. Chem. Soc. **54,** 1499 (1932).
[73] P. Nylén, Tids. Kjemi Bergvesen **18,** 48 (1938).
[74] W. R. Dunstan and E. Goulding, J. Chem. Soc. (London) **75,** 1004 (1899).
[75] J. Meisenheimer, Liebigs Ann. Chem. **397,** 273 (1913).
[76] L. G. S. Brooker and R. H. Sprague, J. Am. Chem. Soc. **63,** 3214 (1941).
[77] L. G. S. Brooker, Rev. Modern Phys. **14,** 275 (1942), and other papers by Brooker and coworkers.
[78] W. Gordy, J. Chem. Phys. **7,** 93 (1939).
[79] W. Gordy and S. C. Stanford, J. Chem. Phys. **8,** 170 (1940); also numerous other papers by Gordy and collaborators.
[80] M. J. Copley, G. F. Zellhoefer, and C. S. Marvel, J. Am. Chem. Soc. **60,** 2666 (1938).
[81] D. Vorländer, Ber. deut. chem. Ges. **36,** 1485 (1903).
[82] D. Vorländer, with J. Fischer and F. Wildner, Ber. deut. chem. Ges. **66,** 1789 (1933).
[83] D. Vorländer, Ber. deut. chem. Ges. **67,** 145 (1934).
[84] R. Dietzel and W. Paul, Arch. Pharm. **276,** 507 (1935); **276,** 408 (1938).
[85] G. Ya. Khalt, Farmatsiya **8,** 26 (1945). For an abstract of the article, see Chem. Abstracts **41,** 2206 (1947).
[86] E. M. Trautner and F. H. Shaw, Australian Chem. Inst. J. & Proc. **12,** 232, 405 (1945).
[87] E. M. Trautner and C. E. Neufeld, Australian Chem. Inst. J. & Proc. **13,** 70 (1946).
[88] O. Folin and A. H. Wentworth, J. Biol. Chem. **7,** 421 (1910).
[89] O. Folin and F. F. Flanders, J. Am. Chem. Soc. **33,** 1622 (1911); **34,** 774 (1912).
[90] O. Folin and F. F. Flanders, J. Biol. Chem. **11,** 257 (1912).
[91] Marion Maclean Davis and Priscilla J. Schuhmann, National Bureau of Standards. Unpublished data.
[92] T. F. Lavine and G. Toennies, J. Biol. Chem. **101,** 727 (1933).
[93] P. Debye and H. Sack, Tables Annuelles de Constantes et Données Numériques **XI,** Section 22 (Paris, 1937).
[94] L. Hartshorn and A. G. Milligan, Int. Crit. Tables VI, 83 (National Research Council, Washington, D. C., 1929).

WASHINGTON, May 5, 1947.

U. S. Department of Commerce
National Bureau of Standards

Research Paper RP1826
Volume 39, September 1947

Part of the Journal of Research of the National Bureau of Standards

Observations on the Control of Grain Size in Magnesium Casting Alloys

By Vernon C. F. Holm and Alexander I. Krynitsky

The methods investigated for the control of grain size in magnesium casting alloys were (1) stirring of carbonaceous solids into the molten alloy, (2) bubbling of carbon monoxide through the melt, (3) melting in a silicon carbide crucible, and (4) treatment with lump magnesite. In the last method, the magnesite was placed deep in the molten metal, where it dissociated thermally, generating a carbonaceous gas that bubbled through the melt. The grain size and tensile properties were determined on specimens from each melt, and these were compared with data from melts that had been superheated as well as those that had received no special treatment. The results indicate that magnesium alloys that have been properly treated with carbonaceous materials possessed grain-size and tensile properties equivalent to those obtained in superheated metal. The magnesite method is a simple, inexpensive, and convenient means of developing a fine-grained structure in magnesium casting alloys.

I. Introduction

It is well known that castings of magnesium alloys tend to have a coarse-grained structure unless suitable steps are taken to insure the development of fine grains. The usual method for producing fine-grained structure has been to superheat the molten metal, a procedure initially described in 1931 in a British patent [1].[1] Superheating, in general, consists in heating a molten magnesium alloy to 250° or 300° C above the liquidus, and after holding it there for a required period, cooling it to the desired temperature and pouring without delay. In recent years, with the increasing interest in magnesium casting alloys, other procedures have been proposed to accomplish this end and thereby avoid some of the disadvantages of superheating. Most of these involve treatment of the molten metal with carbon or a carbon compound. Natural gas [2, 4], acetylene [2, 4], carbon monoxide [5], carbon dioxide [2, 5, 6], and a combination of chlorine and carbon tetrachloride [2, 3] have been recommended for bubbling through the molten metal. Mechanical agitation [4] alone and with additions of finely divided solid carbonaceous matter [6] like coal, peat, graphite, coke, etc., has also been proposed. A process involving noncarbonaceous materials is the German "Elfinal" [7] treatment of the molten magnesium with anhydrous ferric chloride.

Although numerous workers have investigated the problem of developing a fine-grained structure in magnesium castings, and many procedures have been recommended for the control of grain size, knowledge of the actual mechanism involved is far from complete. The fact that superheating and carbon treatments produce similar results and that both are effective only if aluminum is present have suggested that a compound containing aluminum or aluminum and carbon may be essential. Actually, aluminum carbide has been used for controlling grain size [2]. It appears probable that during the treatment some slightly soluble substance is absorbed, and precipitates as very fine particles throughout the melt in the early stages of solidification, to form the necessary nuclei. However, the identity of this substance has not been established.

During the course of an investigation at the National Bureau of Standards of the influence of melting and casting conditions on the properties of magnesium castings, a procedure for grain-size control was devised that appears promising be-

[1] Figures in brackets indicate the literature references at the end of this paper.

cause of its simplicity. The procedure consists in treating the molten alloy with magnesite in lump form. In this paper a comparison is made of the results obtained when magnesium-alloy melts are superheated, treated with solid carbonaceous materials, with carbon monoxide, or with lump magnesite.

II. Equipment and Procedure

Most of the melts were made in a high-frequency induction furnace (960 cycles). The crucibles were of cast steel, 13 in. high and 10½ in. in outside diameter, with a wall thickness of ½ in. The use of steel crucibles diminished the usual "coning up" of the molten charge in the high-frequency furnace and permitted the maintenance of an adequate flux cover on the melt. For comparison, a few melts were made in a gas-fired furnace with a steel crucible, and one special heat was made in a silicon carbide crucible in the high-frequency furnace.

Temperature measurements were made with an 18-gage Chromel-Alumel thermocouple, encased in a closed-end steel tube with an outside diameter of 5/16 in. and a wall thickness of 0.025 in. The melting procedure was as follows:

A small quantity of flux[2] was placed in the cold crucible, the metal charge was introduced, and additional flux was added as required during the melting. At approximately 730° C, the molten metal was stirred and the surface dross was removed. In the superheating treatment flux was added, and the temperature was raised to 900° C. This temperature was maintained for 15 minutes, after which the crucible was removed from the furnace and cooled to the pouring temperature (760° C).

In the other treatments, the additions were made in the range from 730° to 750° C, and flux was added when necessary to avoid excessive oxidation. The temperature was than increased to between 760° and 800° C, after which the melt was cooled to the pouring temperature. All castings were poured at 760° C, and during this operation the stream of metal was protected by the use of a dusting flux consisting of powdered boric acid and sulfur.

[2] A commercial produce known as D-310 of the following composition $MgCl_2$, 50%; KCl, 20%; CaF_2, 15%; MgO, 15%.

The castings were made in green sand molds prepared from a sand formula suggested by Eastwood [8]. The mixture was as follows:

	Percent
Washed silica sand (AFA fineness #80)	91
Western bentonite	4
Powdered sulfur	3
Powdered boric acid	1
Ethylene glycol	1

After thorough mixing, tempering with 2.5 percent of water, and mulling, this sand had the following average properties:

Hardness	84
Permeability	67
Green compression strength	8.8 lb/in

The mold surfaces were generally dusted lightly with powdered sulfur before the molds were closed.

The castings used for grain-size evaluation and for measurement of the tensile properties were tensile test bars of two types. The first were ½-in.-diameter cylindrical bars that were subsequently machined into standard 0.505-in.-diameter tensile test bars. Bars of the second type were cast to this shape and dimension and were tested without machining.

Grain-size determinations were made on a transverse section of a test bar from each heat. To outline the primary crystals, the specimens were held at 385° C for 1 hour, cooled in an air blast, and etched with acetic picral etchant.[3] Tests showed that the slight difference in diameter of the two types of test bars as cast did not affect the grain size appreciably.

The estimation of grain size was accomplished by comparing the microstructure at 100 magnifications with a chart [9] of structures with different average grain diameters. In the extremely coarse structures the approximate grain size was measured under low-power magnification.

The tests were performed on ASTM Alloy No. 4, a magnesium alloy widely used for general-purpose sand-casting. The nominal composition of this alloy is as follows: Aluminum, 6 percent; zinc, 3 percent; manganese, 0.2 percent; balance, magnesium.

[3] 100 ml of a freshly prepared solution of saturated picric acid in 95 percent alcohol, to which is added 10 ml of glacial acetic acid.

III. Results

1. Treatments for Grain-Size Control

Two melts were made to illustrate the coarse grain developed in the absence of any special treatment. In these runs, the 25-lb melts were stirred, skimmed, and heated to about 785° C, cooled to 760° C, and poured. When the melt was made in the high-frequency induction furnace (melt 1, table 1) the cast test bars had an average grain size of 0.05 in., whereas in a duplicate melt (2) made in a gas-fired furnace, the average grain was 0.02 in. in diameter. It is possible that the finer grain of the latter may have been caused by absorption of carbonaceous gas from the furnace. However, the ultimate tensile strengths, 20,900 and 23,100 lb/in.^{2}, are both characteristic of coarse-grained castings.

The effect of superheating to 900° C is shown in the results of melts 3, 4, and 5 recorded in table 1. Superheating produced a uniformly fine grain of 0.004 and 0.005-in. in diameter, regardless of whether the melt was made in the gas or induction furnace or whether the charge consisted of virgin or remelted metal. The results of the tensile tests illustrate the tendency toward somewhat higher values on cast-to-size bars over those that were machined. The ultimate tensile strengths for the machined bars from melts 3, 4, and 5 were nearly all in the range 26,000 to 27,000 lb/in.^{2}, whereas the values for the cast-to-size bars of melt 5 were approximately 3,000 lb/in.^{2} higher. The higher values obtained on the cast-to-size bars are probably due to the presence of a fine-grained surface layer or skin that is removed in the preparation of the machined specimens.

A study of the effectiveness of carbonaceous additions to the metal was made in melts 6 to 10. The additions were made at 730° to 750° C. after which the melts were heated to 800° C and then cooled to the pouring temperature, as recommended by Mahoney, Tarr, and LeGrand [6]. The additions of from 0.2 to 0.8 percent of graphite to the melt were stirred in vigorously for 3 minutes. Small additions of flux were necessary to avoid burning in this operation. Neither the granular graphite (10 to 20 mesh) nor the powdered graphite (finer than 100 mesh) had the desired effect on the grain size of castings from melts 6 and 7 (average

TABLE 1.—*Effect of various treatments of the molten metal on the grain size of the magnesium casting alloy, ASTM No. 4*

Melt No.	Treatment	Grain size (average diameter)	Type of tensile bar	Ultimate tensile strength	Elongation	Remarks
		in.		*lb./in.^{2}*	*Percent*	
1	None	0.050	Machined	20,900	3½	
2	do	.020	do	23,100	4	Gas furnace.
3	Superheated	.005	do	26,500	5	
4	do	.004	do	27,600	5	Gas furnace.
5	do	.005	do	26,300	4½	Charge of remelt metal.
5	do	.005	Cast-to-size	30,100	6	
6	Granular graphite	.030	Machined	21,800	3	
7	Powdered graphite	.030				
8	Granular graphite	.014	Machined	23,900	4	Gas furnace.
9	Silicon carbide crucible	.006				Coarse grain, remelt metal.
10	Carbon monoxide	.005	Machined	27,800	5½	
11	Magnesite	.005	do	25,000	3	
12	do	.005	do	24,900	3	
13	do	.005	do	27,100	3½	
14	do	.005	do	26,100	3½	
15	do	.006	do	23,900	3	
16	do	.005	do	26,300	3½	Coarse grain, remelt metal.
17	do	.004	Cast-to-size	29,800	6	Miscellaneous remelt metal.
18	do	.004	do	30,600	6½	Do.
19	do	.006	do	29,000	5½	Do.
20	do	.005	do	27,900	5	Do.
21	do	.005	do	29,000	5½	Do.
22	do	.005	do	30,200	6½	Do.
23	None	.040	do			Remelt of castings from melt 3.
24	do	.025	do			Remelt of castings from melt 10.
25	do	.070	do	23,000	4	Remelt of castings from melt 22.

diameter, 0.030 in.). In contrast to these melts made in the induction furnace, the addition of granular graphite to melt 8 in a gas-fired furnace, produced a smaller grain size (0.014 in.). This again may be an effect of furnace atmosphere as suggested previously for melt 2. It is believed that the relatively large grain size of the castings from melts 6, 7, and 8 was caused by inadequate contact between the molten alloy and the graphite particles, a condition that might be aggravated by the presence of liquid flux during the treatment.

Another procedure that has been suggested for obtaining better contact between the molten magnesium alloy and a carbon compound consists in making the melt in a silicon carbide crucible. This was tried in melt 9, and the resulting castings indicated a grain size of 0.006 in. This procedure might be promising, except that the flux tends to penetrate and weaken the walls of the crucible.

The effectiveness of carbon monoxide treatment was demonstrated in melt 10, in which the gas was bubbled through the molten metal for 5 minutes at about 750° C. This treatment produced a grain size of 0.005 in., and the tensile properties were equal to those obtained from superheated metal. These results are in agreement with those obtained by Nelson and Holdeman [5]. Apparently some of the carbon monoxide is reduced by the hot magnesium to form nascent carbon, the distribution of which throughout the melt is aided by the agitation produced by the gas bubbles. Carbon dioxide has been shown [2, 5, 6] to be effective in the same manner. This gas can be generated by placing lumps of magnesite in the melt, thus eliminating the need for any special apparatus for the gas treatment.

The use of magnesite as a source of carbon for controlling grain size in magnesium castings was reported by Mahoney, Tarr, and LeGrand [6], who apparently used it in the same manner as other carbonaceous materials, that is, by stirring the powder into the melt. Natural lump magnesite, which has a specific gravity between 2.8 and 3.1, decomposes in the range 200° to 750° C [10] and evolves carbon dioxide.

In order to evaluate the usefulness of lump magnesite for the control of grain size, white California magnesite was used in a series of experiments. The regular melting practice was followed using a steel crucible in the high-frequency furnace.

The results for grain-size control through the use of magnesite are illustrated in melts 11 to 22. In melt 11, approximately 75 g of magnesite, pea size and smaller, was added, but these small pieces remained on the top surface of the melt without the desired action. When a lump about 1 in. in diameter was added it sank into the melt at once and evolved carbon dioxide, which bubbled up to the surface. A thin flux cover was applied, and the decomposition of the magnesite continued for several minutes. The castings were poured at the usual temperature. Subsequent tests indicated an average grain size of 0.005 in., and tensile properties equivalent to those obtained by superheating.

Additional experiments showed that in some instances the gas evolution tended to float the magnesite, thus limiting the effectiveness of the gassing. This problem was solved by placing the magnesite, generally pieces about ¾ in. in diameter, in a perforated sheet-iron container that could be held at the bottom of the crucible. Fifty grams of magnesite was sufficient for the control of grain size in melts ranging from 18 lb. to the capacity of the crucible, 45 lb., and was equivalent to 0.25 to 0.60 percent of the weight of the charge. Treatment with less than 25 g of magnesite was inadequate even for a small charge. The procedure ultimately adopted was to stir and skim the melt at 730° C, and immerse the magnesite for 5 minutes with flux additions as required. The temperature then was raised to 760° C, and the metal was poured.

These experiments indicate that the use of lump magnesite does not introduce large inclusions into the melt, because the decomposition of the magnesite occurs with little change in physical form, and the resulting magnesia is retained in the perforated metal container. Microscopic examination (at×100) of the sections used for grain size evaluation showed no evidence of magnesia inclusions.

Melts 11 to 15 were made with new ingot metal, 16, was a remelt of coarse-grained metal, whereas

mixed stock of both coarse and fine grain was remelted for heats 17 to 22. Satisfactory fine grain was obtained in all cases, and the physical properties of the castings as indicated by tensile tests, were equal to those of superheated metal.

2. Effect of Remelting Fine-grained Castings

Fine-grained magnesium alloys do not always retain a fine structure after remelting. Mahoney, Tarr, and LeGrand [6] pointed out that the need for further grain-size control upon remelting was not as great with metal previously treated with carbon as with superheated ingot or scrap. Eastwood, Davis, and DeHaven [3] state that in their recommended procedure of passing a mixture of carbon tetrachloride and chlorine through the molten alloy, the tendency of the original melt to form fine grains is retained when the castings are remelted. Furthermore, they report that if an alloy is made from 25 percent or more of previously treated metal, the castings will have a fine grain.

A few experiments were included to ascertain the degree to which fine-grain characteristics were retained upon remelting. In melt 23 the charge consisted of metal from melt 3, which possessed an average grain size of 0.005 in., obtained by superheating. When the metal was melted, heated to 760° C, and poured without any special treatment, the resulting grain size in the test casting was about 0.040 in., indicating complete loss of any tendency toward fine grain. Melt 24 was a similar remelt of metal from melt 10, which had been treated with carbon monoxide and which had a grain size of 0.005 in. After remelting and recasting, a grain size of 0.025 in. was obtained. The fine-grained castings from melt 22, which had been treated with magnesite, were remelted in melt 25. Castings from this melt had an extremely coarse grain (0.070 in.). Thus, in each instance the tendency toward fine grain was lost upon remelting, regardless of the method by which fine grains were produced in the original melt.

IV. Discussion of Results

The results verify previous reports on the effectiveness of using gaseous carbon compounds for the control of grain size in magnesium alloy castings. However, treatment with graphite, whether granular or powdered, did not produce the desired fine grain in test castings. The poor showing of this method was undoubtedly due to the difficulty of obtaining a satisfactory contact between the graphite and the molten alloy. The vigorous stirring required during the carbon addition favors oxidation of the metal, and flux additions to control the oxidation appear to limit the action of the carbon. When gaseous carbon compounds are bubbled through molten metal it is thoroughly stirred, and the reducing action of the magnesium apparently produces carbon in a form that can be readily absorbed by the alloy.

The treatment with lump magnesite eliminates the need for the equipment required to bubble gas through a melt and gives the same action and effect as when a gassing treatment is used. The magnesite is inexpensive, can be stored conveniently, and is simple to use. These advantages suggest applications of the method particularly in foundries that operate on a small scale or do not work continuously with magnesium, and where it is desired to obtain fine-grained castings without resorting to superheating or to more elaborate equipment for grain-size control.

The tensile properties of castings produced by the various treatments for grain-size control, shown in table 1, demonstrate the advantage of fine grain, no matter how it is obtained. The highest ultimate tensile strengths were obtained from castings having a grain size of the order of 0.005 in.

The results obtained in remelting fine-grained castings produced from metal that had been superheated, treated with carbon monoxide or with magnesite, indicate that with the technics used in this investigation, the fine-grain characteristics of a magnesium casting are lost when the metal is remelted. The difference between the results for carbon monoxide or magnesite treatments and for those reported elsewhere for carbon tetrachloride–chlorine-treated alloys, is difficult to explain, in view of the fact that the control of grain size in each case has been attributed to the effect of carbon.

The authors gratefully acknowledge the assistance of R. H. Harwell and W. P. Saunders in the experimental work of this investigation.

158728—47——6

V. References

[1] I. G. Farbenindustrie Aktiengesellschaft, British Patent No. 359425, Process for improving the mechanical properties of magnesium alloys (1931).

[2] J. A. Davis, L. W. Eastwood, and James De Haven, Am. Foundryman **8,** 34 (1945); Restricted Reports OPRD, Project No. NRC–546 conducted at Battelle Memorial Institute (1944–45).

[3] L. W. Eastwood, J. A. Davis, and J. C. De Haven, Light Metal Age **4,** 8 (Feb. 1946).

[4] Ralph Hultgren and D. W. Mitchell, Trans. Am. Inst. Mining Met. Engrs. **161,** (Inst. of Metals Div. 323, 1945).

[5] C. E. Nelson and G. E. Holdeman, U. S. Patent No. 2,380,863 (July 31, 1945).

[6] C. H. Mahoney, A. L. Tarr, and P. E. LeGrand, Trans. Am. Inst. Mining and Met. Engrs. 16[illegible] (Inst. of Metals Div. 328, 1945).

[7] The German Magnesium Industry, reported by W. E. McConnica, Technical Intelligence Report, Office of the Chief Quartermaster, p. 5, 41, 67.

[8] L. W. Eastwood, Foundry **73,** 80 (Jan. 1945).

[9] P. F. George, ASTM Bulletin No. 129 (Aug. 1944).

[10] J. W. Mellor, A comprehensive treatise on inorganic and theoretical chemistry, IV, 349 (Longmans, Green, and Co., New York and London, 1934).

WASHINGTON, June 10, 1947.

U. S. Department of Commerce
National Bureau of Standards

Research Paper RP1827
Volume 39, September 1947

Part of the Journal of Research of the National Bureau of Standards

Acceptance Sampling by Variables, With Special Reference to the Case in Which Quality Is Measured by Average or Dispersion[1]

By John H. Curtiss

This paper is devoted to a presentation of the theory and practice of certain types of acceptance sampling plan based on the statistical tests of hypotheses. The basic concepts of the statistical theory are discussed in detail, and are then applied to obtain a number of specific formulas for the single sampling case.

I. Introduction

Acceptance sampling is one of the most interesting and useful applications of modern mathematical statistics. Involving as it does the principles of experimental design and of the testing of statistical hypotheses, it furnishes a proving ground for these theories in which the validity of the basic assumptions is often more readily verifiable, and the correctness of the inference often more rapidly ascertainable, than in many of the standard biometric and sociometric applications.

This paper will be devoted to a presentation of the theory and practice of certain types of sampling plans which are based on statistical tests of hypotheses. Such a plan is intended to act as a sieve for separating unsatisfactory material from acceptable product, to be used with the understanding that there is no guarantee that an occasional poor lot will not accidentally pass through the sieve. The distinction between this type of acceptance plan and certain other popular types is made in the earlier sections of the paper, and the basic concepts of the statistical theory are then discussed in considerable detail. The later sections are devoted to the details of the application of the basic principles to the derivation of explicit formulas for sample size and acceptance numbers. These derivations are given because there seems to be no collection of these derivations and of the formulas themselves readily available in the literature.

[1] An invited address delivered to a joint session of the American Statistical Association and the Institute of Mathematical Statistics in Cleveland, Ohio, on January 27, 1946.

II. What Is Acceptance Sampling?

Because statisticians who are not primarily interested in the engineering applications may wish to follow the discussion for a while, it might be well to start by explaining from the very beginning what acceptance sampling is all about.

Acceptance sampling is a branch of the science of engineering statistics, or industrial statistics, or statistical quality control; that is, the science of dealing with the variability encountered in data arising from engineering experiments and industrial processes. Consider a stream of nominally identical items produced by an industrial process. The items are supposed to be all exactly alike, but of course they never really are. Perhaps the most fundamental and natural way to study their variability is to express the several quality characteristics of each item in terms of a like number of variables which assume numerical values, and then to investigate the properties of the joint frequency distribution of these variables as built up for an aggregate of the items.

One obvious way to perform the investigation consists in actually setting up the frequency distribution in complete detail by making a 100-percent inspection of the items in the aggregate. But for most purposes a much better way to conduct the investigation is to inspect only a portion of the aggregate, and from the data so obtained, make estimates and draw inferences concerning the nature of the frequency distribution of the aggregate. When the portion, or *sample*, is properly chosen, and the methods of inference properly

founded on sound theory, an exact balance can often be achieved between economy in cost of inspection and the needed accuracy of conclusions.

As the name implies, acceptance sampling is a special kind of sampling investigation of an aggregate in which the sample data are used not only to study the distribution of quality characteristics of the aggregate, but also to provide a criterion as to whether the aggregate should be accepted or rejected by a purchaser or consumer.

There are two basic types of acceptance sampling plans. In one type, the entire interest is centered on inferences drawn from the sample relating to the frequency distribution of the aggregate, and rejection or acceptance is specified in accordance with whether or not the inferences so drawn deny or affirm that the frequency distribution of the aggregate meets certain prestated requirements. This will hereinafter be called the *inferential* type of acceptance sampling plan. The other type of acceptance sampling plan is based on the assumption that a rejected aggregate will always be subjected to corrective action (such as 100-percent inspection) which will surely remove all, or a known fraction, of the deficiencies which caused rejection. The plan is then set up so that after the entire inspection process the distribution of the quality-defining variables in the *final outgoing* sequence of aggregates will meet certain prestated requirements. Since the requirements generally pertain to the mean values of the variables in the outgoing sequence of aggregates, this sort of plan might appropriately be termed an *average outgoing quality* type of acceptance sampling plan. Such plans were invented by Dodge and Romig [1].[2]

When there are only a few admissible values of a given quality-defining variable (and the case most frequently encountered is that in which there are only two values, conveniently taken to be 0 and 1, corresponding to whether or not the item meets a specified standard), or when the sampling acceptance-rejection requirements are based on the observed proportion of the total number of sample measurements which fall into a few (say 2 or 3) intervals in the range of a quality defining variable, the associated sampling inspection process is termed acceptance sampling "by attributes." More general cases are grouped together under the generic title of acceptance sampling "by variables." Theoretically, the distinction between "attributes" and "variables" is quite artificial and the boundary is indefinite; "attributes" sampling is really just a highly developed special case of "variables" sampling.

The average outgoing quality type of acceptance sampling plan has been developed only for sampling by attributes and only in the case in which 100-percent inspection (not partial inspection) of rejected lots is specified. The average outgoing quality theory for more general cases, and in particular for the case in which a rejected lot is only partially inspected, apparently has not yet been studied in any detail. This paper will be devoted henceforth entirely to the inferential type of acceptance sampling plan, and the average outgoing quality type will not be mentioned again.

III. An Example

To illustrate how inferential acceptance sampling plans are set up and used, a simple practical example based on a wartime procurement problem will now be presented. For expository purposes and for consideration of commercial confidence, the example has been considerably simplified and the actual dimensions have been altered.

During the war, a small metal cylinder for dispensing insecticide in the form of an aerosol was manufactured in large quantities for the U. S. Navy. The cylinders were manufactured by one set of contractors and filled by another set of contractors. It was necessary to control the volumetric capacity of the cylinder rather carefully, because the filling process used by some of the filling contractors was not automatically adjustable to the size of the individual cylinder. Overfills presented a hazard of explosions when the cylinders were stored in the sun.

A preliminary survey revealed that the frequency distribution of volumetric capacities of cylinders produced by each manufacturer when the manufacturer was aiming at one given nominal capacity, would be of the type illustrated in idealized form in figure 1. The preliminary survey consisted of determining the frequency distribution of a random sample of 1,800 cylinders selected from the first 100,000 produced by each manufacturer. The standard deviation of this observed frequency distribution in each case was about 2.0 cc. A normal (or Gaussian) theoretical distribu-

[2] Figures in brackets indicate the literature references at the end of this paper.

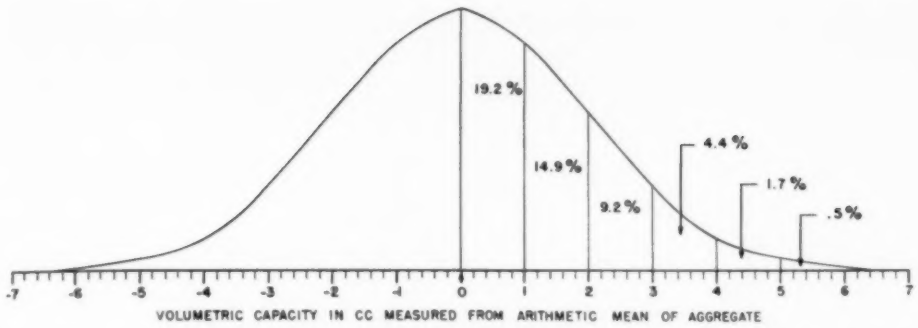


FIGURE 1. *Theoretical frequency distribution of volumetric capacities of cylinders to be used for insecticide dispensers.*

tion gave a satisfactory fit to the data; it is the frequency curve of this distribution which actually appears in figure 1.

The variability in successive determinations of capacity was probably due in no small part to testing error.[3] The chief manufacturing variable affecting volume was in the brazing operation by which the two halves of the cylinder were sealed together. Certain of the manufacturers were known to be experimenting with the nominal size of their cylinders, but it was believed that such changes would probably affect only the mean of the distribution of volumetric capacities, and not the standard deviation.

It seemed probable therefore that the proper control over volumetric capacity could be achieved by first formulating suitable restrictions on the size and pedigree of an aggregate or "lot" of cylinders so that it would have a distribution of the form shown in figure 1, and then imposing requirements on the arithmetic mean of this distribution which were enforced by an acceptance sampling plan.

The appropriate definition of lot will not be detailed here; this is purely an engineering matter. It was decided from the chemical and engineering considerations involved, that if the arithmetic mean of the distribution of volumetric capacities in a lot were to be reduced to less than 530.5 cc, the lot (in effect here, the process) [4] should almost surely be rejected, and if the lot mean remained above 534.0 cc, the lot should almost surely be accepted.

The problem of setting up the acceptance sampling plan that would provide the basis for deciding whether or not a lot meets these requirements can be solved in many different ways. The method which seems to be the most natural one from an intuitive point of view, and which indeed does theoretically have certain optimum properties (see section V below), consists in selecting a random sample of several cylinders from each lot, finding the average of the volumetric capacities of these cylinders, and accepting or rejecting the lot arbitrarily in accordance with whether or not this average is greater than some predetermined "acceptance number."

To determine the sample size and the value of the acceptance number, it is necessary to study the sampling distribution of the average volumetric capacity of a random sample of cylinders taken from a population with a distribution shown in figure 1. It is of course well known that if the population has a normal distribution with mean μ and standard deviation σ, the distribution of the average of a random sample of n observations will be normal with mean μ and standard deviation $\sigma/\sqrt{n}$. The distribution of the average of $n=4$ cylinders is shown in figure 2. From figure 3 it will be seen that if an acceptance number of 532.0 cc is specified, using a random sample of 4, the probability (or "risk") of accidentally rejecting a lot with a mean volumetric capacity of more than 534.0 cc would not exceed 0.023, and the probability of accidentally accepting a lot with mean volumetric capacity of less than 530.5 cc would not exceed 0.067.

[3] It should be noted that the population which is under study in any acceptance sampling plan for a given quality defining variable is the population of *apparent values* of the variable as determined by some test or inspection procedure. Thus the variance of the population is made up of two components: one due to the nonuniformity in the material and one due to lack of precision in the test.

[4] See section XII.

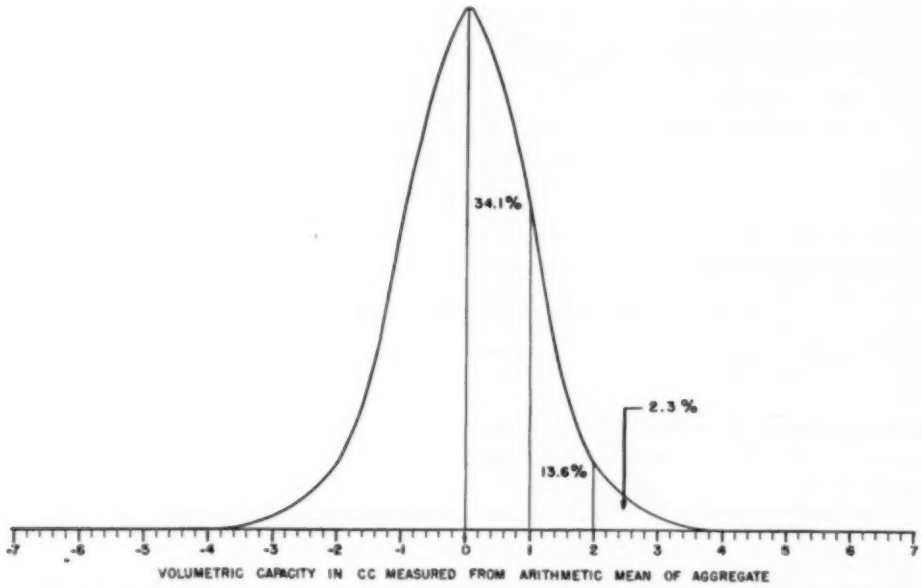


FIGURE 2. *Probability distribution of mean of random sample of four cylinders, derived from the theoretical distribution in figure 1.*

Either one of these two risks can obviously be decreased at the expense of increasing the other by merely changing the position of the acceptance number appropriately. By using a larger sample size, or by multiple or sequential sampling, it would be possible to reduce both of the risks simultaneously. The general order of magnitude of the risks in the sampling plan illustrated by figure 3 is considered to be appropriate for such work.

IV. Acceptance Sampling and Statistical Inference

Inferential acceptance sampling plans may be classified as statistical or nonstatistical, in accordance with whether or not they are based on the principles of statistical inference.

In nonstatistical inferential plans, sample items are purposively selected in accordance with prior information, and from the observations on the sample items a supposedly *certain* inference is made concerning the nature of the sampled aggregate. Such acceptance sampling plans will not be considered in this paper. They are useful and efficient in special cases, but their theory is a function of the application and does not lend itself readily to a general discussion.

Statistical inferential acceptance sampling plans are characterized by a selection of sample items from an aggregate in accordance with some method which insures that the observations on the respective items in the sample have a joint probability distribution which bears some direct and calculable relation to the frequency distribution of the aggregate. Through the use of this probability distribution, an *uncertain* inference into the nature of the distribution of the aggregate is drawn from the data of a sample. The example given in section III was a statistical acceptance sampling plan. The "random" selection of the sample cylinders implies a unique determination of the probability distribution of the sample in terms of the frequency distribution of volumetric capacities in the sampled aggregate.

The general theory of the uncertain inferences involved in statistical acceptance sampling plans may conveniently be identified with a branch of modern mathematical statistics called tests of statistical hypotheses.[5] A statistical hypothesis is a statement concerning the unknown frequency

[5] Simon [2] prefers to identify the theory of statistical acceptance plans with the theory of estimation of population parameters. The approach via the theory of statistical hypotheses, which was pioneered by Dodge and Romig [1], has the advantage that it provides a quantitative basis for a rational selection of the accuracy of estimation required in a given application; see section VI.

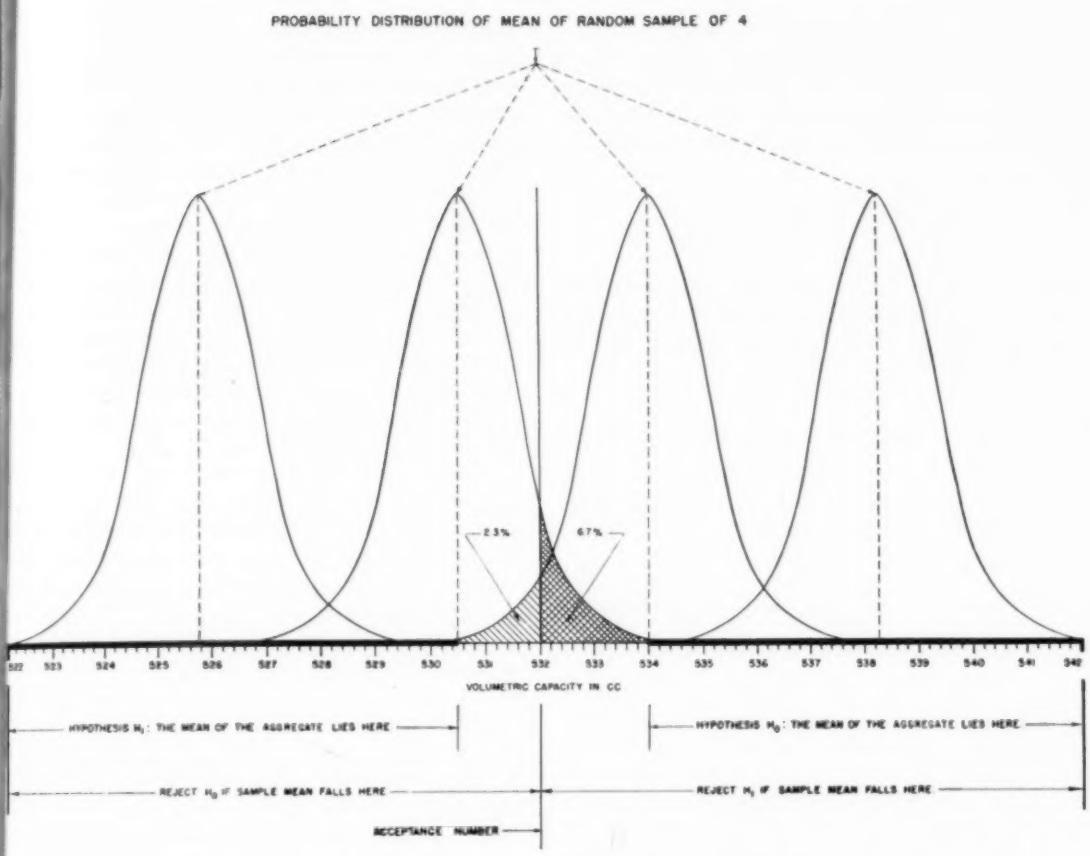

PROBABILITY DISTRIBUTION OF MEAN OF RANDOM SAMPLE OF 4
2.3%
6.7%
522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542
VOLUMETRIC CAPACITY IN CC
HYPOTHESIS H1: THE MEAN OF THE AGGREGATE LIES HERE
HYPOTHESIS H0: THE MEAN OF THE AGGREGATE LIES HERE
REJECT H0 IF SAMPLE MEAN FALLS HERE
REJECT H1 IF SAMPLE MEAN FALLS HERE
ACCEPTANCE NUMBER

FIGURE 3. *Test of hypothesis concerning the arithmetic mean of the aggregate.*

distribution (or probability distribution) of a variate X (which may be a vector with several components), to the effect that this unknown frequency distribution belong to a certain well-defined subclass ω of a class Ω of admissible frequency distributions of X. The hypothesis is tested by first drawing a sample of observations on the variable X, say $(X_1, X_2, \ldots, X_n)$, in such a way that the probability distribution of the sample can be calculated from the distribution of X, and then noting just where the point defined by the coordinates $(X_1, X_2, \ldots, X_n)$ falls in n-dimensional Euclidean space. If this sample point falls on a certain predetermined region w of the n-dimensional space, the hypothesis is rejected. The subset w is called the "critical region"[6] for the test.

In the application to acceptance sampling, the variable X is a measure of the quality of the individual item, and the hypothesis to be tested is that the unknown frequency distribution of X in the sampled aggregate of items belongs to a certain subclass of distributions which is identified with definitely "good" or acceptable quality (alternatively, "bad" or unacceptable quality) with respect to the characteristics measured by X. Thus in the example of section III, X is volumetric capacity; the admissible class Ω of frequency distributions consists of all normal distributions with standard deviations equal to 2.0 cc; the subclass ω representing acceptable quality consists of all such frequency distributions with means not less than 534.0 cc; and the purpose of the sampling plan can be considered to be to test

[6] For the benefit of mathematicians reading this paper, it might be observed that this usage of the word "region" is at variance with the usage in classical mathematical analysis; the critical region is not necessarily an open connected set.

the hypotheses (designated as H_0 in fig. 3) that the distribution of X lies in this subclass.

The critical region for this test, chosen on intuitive grounds in section III, was the region in the sample space of four dimensions defined by the inequality

$$\frac{X_1+X_2+X_3+X_4}{4}<532.0, \qquad (1)$$

where X_1, X_2, X_3, and X_4 represent four observations on X. In general, in acceptance sampling whenever the critical region is defined by an inequality such as

$$c_1<f(X_1, X_2, \ldots, X_n)<c_2,$$

where f is some continuous function of the apparent variables, the numbers c_1 and c_2 can appropriately be called *acceptance numbers*. It is also natural to call the critical region the *region of rejection* in acceptance sampling work.

V. Choice of the Optimum Type of Critical Region

It is fairly clear that the theory of tests of statistical hypotheses must revolve to a considerable extent around the choice of the critical region. Such a region can always be chosen in infinitely many ways so that the probability that the sample point falls in it, given that the hypothesis under test is true, is not greater than a given arbitrarily small number. For example, the critical region used in section III is the region defined by the inequality (1), and the maximum probability of the sample point falling into it under the hypothesis of good quality is 0.023. But another critical region for which this maximum probability is also 0.023 is defined by the inequality

$$532.9<\frac{X_1+X_2+X_3+X_4}{4}<533.0, \qquad (2)$$

and the reader will have no difficulty in concocting other critical regions with the same maximum probability of rejection. Which one is the "best"?

To answer this sort of question, Neyman and Pearson [7] introduced the idea of the "power" of a test of a statistical hypothesis. This concept is based on the obvious fact that the probability that the sample point falls in a predetermined critical region w depends on what the unknown frequency distribution of X really is. That is, for each member of the class Ω of admissible frequency distributions of X, there is a probability, theoretically calculable, that the sample point falls in w. The power of the test is simply defined to be this probability, and is thus a functional depending on the distribution of X (and also on w). This means that if the frequency distribution of X really belongs to the subclass ω (that is, the hypothesis is really true), then the power of the test is the probability of (erroneously) rejecting the hypothesis, and if the frequency distribution of X does not belong to ω, then the power of the test is the probability of (correctly) rejecting the hypothesis.

Common sense indicates the right way to choose the critical region for a test is to choose it in such a way that when the hypothesis is true, the power is as small as possible, and when the hypothesis is not true, the power is as large as possible. A more careful formulation of this principle would proceed as follows: For any given fixed sample size n,[8] consider all possible critical regions for which the power is less than or equal to a given number α when the hypothesis is true. From these, select the particular region which has the property that for each frequency distribution not in ω it makes the power greater than do all the other regions. Such a region is said to provide a uniformly most powerful test. When this sort of optimum region does not exist, various compromises have been worked out which rationalize and systematize intuitional solutions of the problem. In some cases where a critical region which is known to be in some sense optimum does exist, the boundary is so complicated that for practical reasons it is desirable to use a simpler critical region. An example of such a case will be discussed in section VIII.

To return to the example of section III, the power of the test for any member of the class Ω and for any critical region defined by an inequality like (1) or (2) is very easily calculated by simply noting that the probability distribution of $(X_1+X_2+X_3+X_4)/4$ has exactly the same mean as the parent frequency distribution in Ω, and has the standard deviation $2/\sqrt{4}=1$ cc. Therefore, the probability that the sample observations fall so

[7] See, for example, Wilks [3], chapter VII.

[8] If sequential sampling is used, this clause can be omitted (see Wald [6]). If multiple sampling is used, the clause should be replaced by a specification of the sample size at each step of the sampling procedure.

that they satisfy an inequality such as (1) or (2) can quickly be looked up in a table of a normal distribution. It appears reasonable from figure 3, which shows the distribution of $(X_1+X_2+X_3+X_4)/4$ drawn for four typical members of the class Ω, that among all critical regions of the type given by inequalities such as (1) and (2), the one given by (1) gives the uniformly most powerful test in the sense above defined. This can be proved rigorously with little trouble. It is not so easy to show that this same critical region gives the uniformly most powerful test among all possible imaginable critical regions in space of four dimensions, and not just among those critical regions defined by inequalities to be satisfied by the sample mean, but it can be proved that such is indeed the case [5].

VI. Adjustment of the Power of the Test

Presuming that an optimum (or at least, satisfactory) type of critical region for a given statistical hypothesis is known, the chief remaining question in setting up the test, or acceptance sampling plan, is that of arranging for the appropriate power. It was previously pointed out that the power of the test depended on the distribution of the variate X to which the hypothesis applied, and on the shape and size of the critical region w. In addition, it should now be observed that for a fixed specification of the distribution of X and for a fixed w, the power also depends on the probability distribution of the sample point $(X_1, X_2, \ldots, X_n)$, which is determined by such things as method of selection of the sample and the sample size.

The problem of deciding upon the appropriate power for a test is complicated by the fact that in most cases of practical importance, given any fixed critical region, the power functional does not have a sharp discontinuity as the frequency distribution of X crosses the boundary between ω and the remainder of Ω. In other words, if the hypothesis is almost, but not quite, true, the probability that the sample point falls into the critical region is about the same as if the hypothesis was almost, but not quite, untrue. This phenomenon can be seen very easily in figure 3. Obviously as the mean of the true distribution moves continuously leftward from the value 534.0 cc, which represents the boundary of good quality, the shaded area to the left of the point 532.0 cc (which represents the power of the test) continuously increases.

A convenient way of getting around this difficulty is to recognize that in the applications there is often no real need for distinguishing sharply between a case where the hypothesis is true and another case where it is "just a little bit" untrue. In many situations it is possible from practical considerations to select a second well-defined subclass ω_1 of the class Ω of admissible frequency distributions, such that there is a positive distance (as defined in some reasonable way) between ω and ω_1, and such that from the viewpoint of the application, ω_1, and ω are subclasses that are so radically different that if the true distribution of X really belongs to ω_1, then there should be a high probability of rejecting the hypothesis that it belongs to ω. It is natural in such a case to introduce a second statistical hypothesis into the picture, consisting of a statement that the unknown frequency distribution of X belongs to the subclass ω_1. The test is then conveniently thought of as affording a criterion of choice between the two hypotheses.

This approach, which will henceforth be called the method of alternative hypotheses, leads to simple quantitative rules for determining the appropriate power, as for example that the power shall not be greater than a preassigned number α when the distribution of X is truly a member of ω, and the power shall not be less than a preassigned number $1-\beta$ when the distribution of X is a member of ω_1. For a given method of sample selection and a given type of critical region, this sort of requirement will in many practical cases uniquely determine a minimum sample size and a corresponding position for the critical region, as will be seen in section VIII.

In the application of the theory of statistical tests to acceptance sampling, the method of alternative hypothesis is particularly useful and convenient. It is seldom indeed that the boundary between the characterization of a "good" lot and a "bad" one is sharply defined; but given a class of admissible frequency distribution of a quality-defining variable X, it is often possible to pick out two isolated subclasses which respectively represent definitely "good" or definitely "bad" quality with respect to the characteristics measured by X. When this has been done, the maximum [9] probability of erroneously rejecting an aggregate that is really in the good subclass is

[9] Maximum is here used in the sense of least upper bound.

called the *producer's risk*, and the maximum (see footnote 9) probability of erroneously accepting an aggregate that is really in the bad subclass is called the *consumer's risk*. In the terminology of the preceding paragraph, if β represents good quality and ω_1 represents bad quality, then α is the producer's risk and β is the consumer's risk—a notation that will be adhered to in the sequel.

The example of section III illustrates the method of alternative hypotheses. It was pointed out previously in section IV that the admissible class Ω of frequency distributions in this example consists of all normal distributions with standard deviations equal to 2.0 cc, and the subclass ω, identified with a good quality, consists of all frequency distributions in Ω with means not less than 534.0 cc. A second subclass ω_1, identified with unacceptable quality, is also defined in the example; this ω_1 subclass consists of the members of Ω with means less than or equal to 530.5 cc. The producer's risk of the sampling plan thereupon chosen is 0.023, and the consumer's risk is 0.067.

VII. Parametric Case and Operating Characteristic

The study of the power of a statistical test is considerably simplified when the members of the class Ω of admissible frequency distributions are uniquely determined by the values of one or more parameters. The discussion in the preceding sections has not been limited to the parametric case because the basic definitions and concepts can be stated just as easily for the general case as for the parametric case. Moreover, work is proceeding rapidly now on the theory of nonparametric tests, and it would seem to be worth while to lay down general definitions here with an eye to the day when nonparametric tests come into general use in acceptance sampling. But parametric hypotheses are still used far more commonly in the applications than nonparametric ones; this is especially true in acceptance sampling applications.

The case illustrated in the example of section III, where each member of the class Ω was uniquely determined by the value of its mean, is an example of the important case in which the class Ω is a one parameter family. Let the parameter in a one parameter family of admissible frequency distributions be denoted by θ. For a fixed critical region w, and method of sampling, the power of the test is a single-valued function of θ, which can be represented graphically in the usual way. In the theory of statistical inference, this graph is called the *power curve* of the test; in acceptance sampling work, it is called the *operating characteristic*. The power curve or operating characteristic of the sampling plan of section III will be found in figure 4.

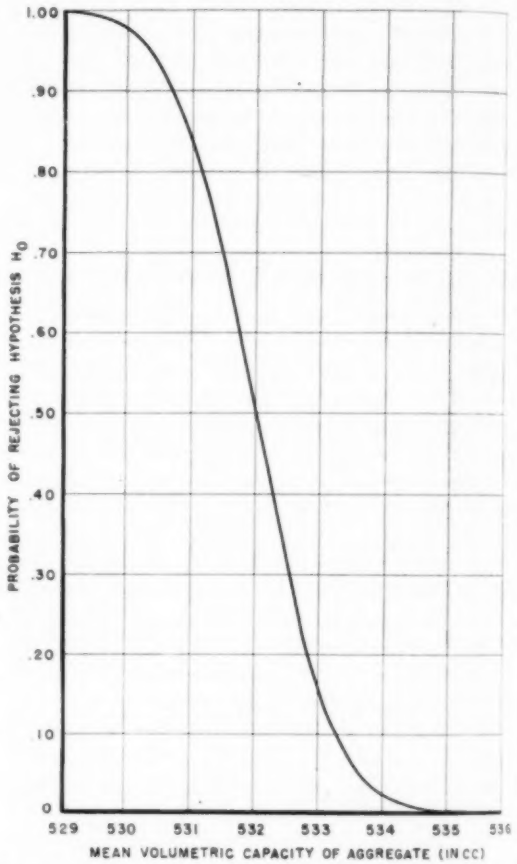


FIGURE 4. *Operating characteristic of the sampling plan of figure 3.*

By drawing the power curves, or operating characteristics of a number of proposed tests, corresponding to different positions of the critical region and different sample sizes (and perhaps even to different methods of sampling), a rational choice of the most suitable test can often be made.[10]

[10] Collections of operating characteristics for the attributes case will be found in [6] and in [7]. A collection of power curves for certain standard tests of statistical hypotheses for sampling by variables is given in [8].

The choice can be based on one or more of a number of rather obvious properties of the various curves, such as their relative slopes and their heights at specially chosen values of θ. The alternative hypothesis approach described in considerable generality above can also be used profitably in conjunction with a power curve. The subclasses ω and ω_1 are respectively represented by intervals (or more general point sets) on the θ-axis of the power curve, and from inspection of the curve the maximum probabilities of rejecting the alternative hypotheses when θ lies in these intervals can easily be estimated.

VIII. Sample Size and Region of Rejection: Normal Distribution

In the present section, and in the following one, the methods of alternative hypotheses will be applied to obtain some general formulas of practical importance for sample sizes and regions of rejection of acceptance sampling plans involving parametric hypotheses. The parameters to which the hypotheses pertain will be the mean and the standard deviation of the frequency distribution of a quality-defining variable X in the sampled aggregate. Single sampling only will be considered; this means that a sample of n items is to be drawn from the aggregate, the value of X for each of the n items is to be observed, and on the basis of these observations the aggregate is to be sentenced. It will further be assumed that the sample is so selected that the observations on the items are statistically independent, and the probability distribution of each observation is identical with the unknown frequency distribution of X. This can be accomplished, for example, by what is generally known as "random" selection from a finite aggregate with replacement after each drawing. The assumption will not be strictly valid if the sample items are selected "at random" *without* replacements, but in the practical applications of sampling by variables the size of the aggregate is usually so much larger than that of the sample that the requirement of replacements can be dropped without causing any appreciable change in the probability distribution of the sample. As a convenient rule of thumb, it might be stated that the formulas to be derived are valid for sampling without replacements if the ratio of aggregate size to sample size is greater than 10 to 1.

Consider first the case in which the parameter to which the hypotheses are to be applied is the mean μ of the frequency distribution of the aggregate, as in section III. Suppose further that this distribution is normal, that the value of its standard deviation σ is known, say $\sigma=\sigma_0$, and that two values of μ, μ_0, and μ_1 can be found, so that acceptable quality is identifiable with μ lying in the interval $\mu_0 \leqq \mu < \infty$, and unacceptable quality is identified with $-\infty < \mu \leqq \mu_1$. (This is the case exemplified in sec. III.) As pointed out in section V, a critical region of the type $\bar{x} < c$ will provide a uniformly most powerful test; here c is some constant and

$$\bar{x} = \frac{X_1 + X_2 + \ . \ . \ . \ + X_n}{n},$$

where X_1, X_2, . . ., X_n, denote the observations on the sample items. Let α be the producer's risk and β be the consumer's risk. Then the definition of these risks yields the following pair of simultaneous inequalities

$$\text{Prob.}\ [\bar{x} < c \mid \mu_0 \leqq \mu < \infty,\ \sigma = \sigma_0] \leqq \alpha, \tag{3}$$

$$\text{Prob.}\ [\bar{x} \geqq c \mid -\infty < \mu \leqq \mu,\ \sigma = \sigma_0] \leqq \beta. \tag{4}$$

(The first formula is to be interpreted as stating that the probability of $\bar{x} < c$, given that $\mu_0 \leqq \mu < \infty$ and $\sigma = \sigma_0$ is less than or equal to α, and the second formula has an analogous meaning.)

Now it can be seen, say from a figure like figure 3, that for a fixed c, if μ is allowed to approach μ_0 from the right, while the distribution of $\bar{x}$ is otherwise kept unchanged, then Prob. $[\bar{x} < c]$ continually increases, and achieves a maximum at $\mu = \mu_0$. Similarly, Prob. $[\bar{x} > c]$ achieves a maximum for all $\mu \leqq \mu_1$ when $\mu = \mu_1$. Thus (3) and (4) can be revised to read

$$\text{Prob.}\ [\bar{x} < c \mid \mu = \mu_0,\ \sigma = \sigma_0] = \alpha. \tag{5}$$

$$\text{Prob.}\ [x \geqq c \mid \mu = \mu_1,\ \sigma = \sigma_0] = \beta. \tag{6}$$

As $\bar{x}$ has a normal distribution with mean μ and standard deviation $\sigma_0/\sqrt{n}$ it follows that $(\bar{x}-\mu)/(\sigma_0/\sqrt{n})$ has a normal distribution with mean zero and unit standard deviation. Tables of this sort of normal distribution (henceforth to be called a standard normal distribution) are

widely available. Consider (6) first. It follows that

$$\text{Prob.}\,[\bar{x} \geqq c | \mu = \mu_1, \sigma = \sigma_0]$$

$$= \text{Prob.}\left[\frac{\bar{x}-\mu_1}{\sigma_0/\sqrt{n}} \geqq \frac{c-\mu_1}{\sigma_0/\sqrt{n}} \middle| \mu = \mu_1, \sigma = \sigma_0\right]$$

$$= \left[\text{Area under standard normal curve from ordinate at } \frac{c-\mu_1}{\sigma_0/\sqrt{n}} \text{ to } +\infty\right]. \quad (7)$$

Let k_p be the $100p$ percent point on the standard normal scale, that is, the point at which the ordinate of the standard normal curve bounds a tail area exactly equal to p. Then combining (6) and (7),

$$\frac{c-\mu_1}{\sigma_0/\sqrt{n}} = k_\beta. \quad (8)$$

A similar argument applied to (5) leads to the equation

$$\frac{c-\mu_0}{\sigma_0/\sqrt{n}} = -k_\alpha. \quad (9)$$

The solution of (8) and (9) simultaneously for n and c yields formulas for these unknowns in terms of σ_0, k_α, and k_β, which provide a general solution to the problem of setting up the sampling plan. These formulas will be found under case 1 in the list of formulas in the appendix. Analogous formulas for the case in which the two hypotheses are interchanged are given there under case 2.

A slight generalization is obtained by remarking that the argument that led from (3) and (4) to (5) and (6) is valid if the class of admissible distributions is extended from just those normal distributions with $\sigma = \sigma_0$ for all values of μ, to include all normal distributions with $\sigma \leqq \sigma_0$ when μ lies in the interval identified with good quality, and with $\sigma \leqq \sigma_1$ when μ lies in the interval identified with bad quality (where σ_1 is any arbitrary positive number), and with no restriction at all on σ when μ lies in the interval identified with neither bad nor good quality.

The region of rejection $\bar{x} < c$ is then no longer a uniformly most powerful critical region, but it does still have certain optimum properties. This generalization is incorporated in the formulas given in the appendix.

The significance in practice of this extension of the class of admissible distributions is that while in a given instance it is seldom known exactly what the standard deviation of the distribution of the aggregate is, frequently a good guess can be made at a reasonable upper bound for this standard deviation. The practical side of making estimates of bounds for dispersion will be considered in more detail in section XIII.

A case which is often encountered is one in which the standard deviation of the distribution of a quality-defining variable X is directly proportional to the mean of X. Minor adjustments in the above argument yield the formulas listed under cases 3 and 4 in the appendix. The region of rejection used in these formulas is not a uniformly most powerful region, and in fact a theoretically "better" region can be found fairly easily.[11] However, this better region is a hypersphere in the sample space of n dimensions with radius and center which depend in a complicated way on μ_0, μ_1, and n. Although the matter has not been carefully investigated, there would seem to be very little loss in power in using a critical region of the type $\bar{x} < c$, as is done in the formulas in the appendix, and the gain in simplicity is very considerable. It would be of interest to make a more thorough study of this point.

Turning now to the case in which the hypotheses are to apply to the standard deviation σ of the sampled aggregate, an optimum type of critical region for testing the hypothesis $\sigma \leqq \sigma_0$ against the hypothesis $\sigma \geqq \sigma_1$ in the case of a normal aggregate $\sigma_0 < \sigma_1$ is known to be given by the inequality $s^2 > c$, where c is a positive constant, and s^2 is the usual unbiased sample estimate of σ^2; that is,

$$s^2 = \frac{\sum_{j=1}^{n}(X_j - \bar{x})^2}{n-1}.$$

Inequalities (3) and (4) become

$$\text{Prob.}\,[s^2 > c | \sigma \leqq \sigma_0] \geqq \alpha \quad (10)$$

$$\text{Prob.}\,[s^2 \leqq c | \sigma \geqq \sigma_1] \leqq \beta \quad (11)$$

Now the probability distribution of $(n-1)s^2/\sigma^2$ depends only on n and not on μ or σ, and is a standard tabulated distribution known as a chi-square distribution with $n-1$ degrees of freedom.[12]

[11] This better critical region is the region for the "likelihood ratio test"; see p. 150 of [3].

[12] See [9], p. 150. For tables of the chi-square distribution, see [9], p. 159.

This means that since

$$\text{Prob. } [s^2 > c] =$$

$$\text{Prob. } \left[\frac{(n-1)s^2}{\sigma^2} > \frac{(n-1)c}{\sigma^2}\right], \tag{12}$$

for a fixed value of n and c, the probabilities in this identity increase as the quantity $(n-1)c/\sigma^2$ decreases; that is, as σ increases. This remark leads to the following revisions of (10) and (11):

$$\text{Prob. } [s^2 > c | \sigma = \sigma_0] = \alpha. \tag{13}$$

$$\text{Prob. } [s^2 \leqq c | \sigma = \sigma_1] = \beta. \tag{14}$$

Now let $\chi^2_{n,\,p}$ denote in general the $100p$ percent point of the chi-square distribution with n degrees of freedom; that is, the point at which the ordinate of the frequency curve of this distribution bounds a *right-hand* tail area exactly equal to p. By applying the identity (12) to (13) and (14), the following simultaneous equations are obtained:

$$\frac{(n-1)c}{\sigma_0^2} = \chi^2_{\alpha,\,n-1},$$

$$\frac{(n-1)c}{\sigma_1^2} = \chi^2_{1-\beta,\,n-1}.$$

The solution of these simultaneous equations for n and c yields the formulas that define the sampling plan. The best way to obtain the solution seems to be as follows:

Replace the equations by the equivalent pair

$$\frac{\chi^2_{\alpha,n-1}}{\chi^2_{1-\beta,n-1}} = \frac{\sigma_1^2}{\sigma_0^2}. \tag{15}$$

$$c = \frac{\sigma_0^2 \chi^2_{\alpha,n-1}}{n-1}. \tag{16}$$

For fixed α and β, the quantity on the left side of (15) decreases as n increases. By inspection of a table of the percentage points of the chi-square distribution,[13] determine the first value of $(n-1)$ which makes the left member of (15) greater than (or equal to) the right member. Then substitute into (16) to obtain c.

Most tables of the χ^2 distribution stop at $n=30$. Beyond this value of n, a satisfactory approximation is obtained by treating s as if it had a normal distribution with mean equal to σ and with standard deviation equal to $\sigma/\sqrt{2n-2}$.

[13] See footnote 12.

The derivation of formulas for n and c then becomes essentially the same as that for case 4 in the appendix; the results are given under case 7.

IX. Sample Size and Region of Rejection: Nonnormal Distribution

The principal limitation to the use of the formulas so far developed is that they involve the assumption that the distribution of the quality-defining variable X in the sampled aggregate is of the normal or Gaussian type. If in any given case, the assumption of normality appears to be inapplicable, it may happen that some other fairly tractable form of distribution can be specified. The general principles outlined in the preceding sections for determining sample sizes and regions of rejection will again be applicable, but rather troublesome computational difficulties may be encountered in determining the distribution of the sample mean and in solving for the sample size and boundary of the optimum critical region.

Moreover, in practice, cases are frequently encountered in which, due to lack of "statistical control" (see section XII) or lack of any historical data concerning the production process, very little *a priori* information is available concerning the possible form of the distribution of the variable X in a given aggregate. A method for dealing with such cases will be described in the remaining paragraphs of this section. It naturally leads to somewhat higher sample sizes than are obtained when for equivalent hypotheses and risks of error the form of the distribution is completely specified in advance. However, the increase in sample size over that obtained even with the assumption of normality is often not unreasonably great, and it is suggested that the method about to be described might be used whenever there is serious doubt about normality, unless there exists *a priori* knowledge of the physical nature of the production process which leads quite definitely to the conclusion that X should have some special form of nonnormal distribution, such as a Poisson distribution.

The method consists in making use of the various inequalities of the so-called Tchebycheff type, which place upper bounds on the proportion of a distribution that can lie at more than a preassigned distance from a central point in the distribution.[14]

[14] See [10] p. 182 to 183.

The best-known and most frequently employed of these inequalities is the Bienaymé-Tchebycheff inequality,[15] which presupposes no restrictions at all on the distribution except that there is a finite standard deviation. As might be expected from the absence of prior information assumed, the use of this inequality to derive formulas analogous to those of section VIII leads to very high sample sizes compared to those obtainable under the assumption of normality. It is possible to improve the Bienaymé-Tchebycheff inequality somewhat by imposing certain additional restrictions on the distribution, such as that it be continuous and unimodal.[16] This type of restriction is not unrealistic in acceptance sampling work.[17] But another and more fundamental difficulty in the application of the Bienaymé-Tchebycheff inequality to acceptance sampling by variables is that the inequality seems to fail to take full advantage of the usually rapid approach of the sampling distributions of $\bar{x}$ and s to the normal form as n increases.

Thus in spite of the popularity of the Bienaymé-Tchebycheff inequality and its various modifications, in the case of hypotheses concerning μ, it would seem that the appropriate type of Tchebycheff inequality to use is one which is known as the inequality of S. Bernstein. This inequality has the double virtue of appearing to avoid the difficulty just now mentioned, and of involving restrictions on the distribution of X that are automatically satisfied in most acceptance sampling work. The Bernstein inequality may be written in the form of a pair of inequalities as follows: [18]

$$\text{Prob. } [\bar{x}-\mu \geqq K] \leqq e^{-nK^2/(2\sigma^2+2hK)}, \qquad (17)$$

$$\text{Prob. } [\bar{x}-\mu \leqq -K] \leqq e^{-nK^2/(2\sigma^2+2hK)}, \qquad (18)$$

and the restrictions on the distribution of the aggregate are that the absolute moments

$$\nu_j = E(|X-\mu|^j),\ j=1, 2, \ldots,$$

all exist and satisfy the inequality

$$\nu_j \leqq \frac{\sigma^2}{2} j! h^{j-2},\ j=2, 3, \ldots,$$

where h is a positive number. These conditions are satisfied if, for example, the distribution of X is bounded, which is usually the case in practice. If the distribution of X lies entirely in the interval $(\mu-\Delta,\ \mu+\Delta)$, then it can be shown that h can be taken as equal to Δ.[19]

The use of the inequalities (17) and (18) to derive formulas for sample size and critical regions for hypotheses relating to μ, follows closely in the pattern set up in section VIII. The same type of regions of rejection will be used as before; these regions at least have optimum properties in the limit as the sample size tends to infinity.

The argument used in section VIII is applicable without change down through (5) and (6). At that point the inequalities (17) and (18) are introduced. The analogue of (7), obtained by using (17), is

$$\text{Prob. } [\bar{x} \geqq c | \mu=\mu_1,\ \sigma=\sigma_0] =$$

$$\text{Prob. } [\bar{x}-\mu_1 \geqq c-\mu_1 | \mu=\mu_1,\ \sigma=\sigma_0] \leqq e^{-\frac{n(c-\mu_1)^2}{2\sigma_0^2+2h(c-\mu_1)}} = \beta,$$

where $e=2.71828\ldots$. That is,

$$\frac{n(c-\mu_1)^2}{\sigma_0+h(c-\mu_1)} = -2 \log_e \beta. \qquad (19)$$

Similarly, applying (18) to (5), the following equation is obtained:

$$\frac{n(c-\mu_0)^2}{\sigma_0^2-h(c-\mu_0)} = -2 \log_e \alpha. \qquad (20)$$

The two simultaneous equations (19) and (20) define n and c. The extension to the case in which the *a priori* information $\sigma=\sigma_0$ is replaced by $\sigma \leqq \sigma_0$ for H_0 and $\sigma \leqq \sigma_1$ for H_1 is again valid. The only change in the equations (19) and (20) necessitated by this generalization consists in replacing σ_0 by σ_1 in (19).

Unfortunately, (19) and (20) happen to be equations of the third degree in the unknowns, and the general solution is somewhat complicated. The situation is greatly simplified from the algebraic point of view by taking $\alpha=\beta$. The solution for this case, and with $\sigma_0=\sigma_1$ for additional simplicity, is given in cases 5 and 6 in the appendix. The reader will have no great difficulty in obtaining the solution of (19) and (20) in more general cases, either by graphical methods or by straightforward algebra.

[15] See [10] p. 183.
[16] See [10] p. 183, also [11].
[17] Simon [2] discusses in some detail the use of the sharpened Bienaymé-Tchebycheff inequality in industrial sampling work.
[18] See [12] p. 204 to 205; also [13].
[19] See [12] p. 205.

X. The Case of Unknown σ

It has doubtless been noticed by the reader that the formulas so far derived for acceptance sampling plans involving hypotheses relating primarily to μ, have unfortunately always explicitly involved upper bounds for σ. This situation stems from the fact that the power curves for any efficient test of a statistical hypothesis concerning μ will in general always depend on higher moments of the distribution of μ, except in the case of certain very special distributions, where σ is a known function of μ. (An example of such a special distribution is the Poisson distribution, where $\sigma=\sqrt{\mu}$.)

If it can be assumed without too much error that the distribution of X is normal, then it is possible to devise a test of a hypothesis relating to μ, which has the property that the power curves corresponding to different values of σ all go through one point, say the point (μ_0, α) or the point (μ_1, β) in the notation of section VIII and the appendix. Moreover, the power curve of the test corresponding to each given value of σ either always rises or always falls as μ increases, so that if the hypothesis to be tested is of the form $\mu_0 \leqq \mu < \infty$, say, and if the test is arranged so that each power curve passes through the point (μ_0, α), then it is possible further to arrange the test so that the power for $\mu > \mu_0$ is always less than α no matter what σ may be.

Translated into the language of acceptance sampling, this all means that if the distribution of X in the sampled aggregate can be assumed to be normal, and if good quality is defined, say, by $\mu_0 \leqq \mu < \infty$ and bad quality by $-\infty < \mu \leqq \mu_1$ (as in case 1 in the appendix), then there is a way to set up a sampling plan that will have a preassigned producer's risk, or alternatively a preassigned consumer's risk, regardless of the standard deviation of the aggregate. But only a chosen one of the two kinds of risks can be preassigned; when a series of aggregates having differing values of σ are sampled, the value of the opposite risk will be different for each aggregate in the series, and will increase with σ.

Assuming (as always) the method of sampling described in section VIII, the critical region (or region of rejection) for the test in question is defined by inequalities of the type

$$\frac{\bar{x}-\mu'}{s/\sqrt{n}} < c, \tag{21}$$

$$\frac{\bar{x}-\mu'}{s/\sqrt{n}} \geqq c, \tag{22}$$

where μ' is the particular value of μ for which it is desired that the power of the test shall be independent of σ. In case the producer's risk is to be fixed for a hypothesis of good quality given by $\mu_0 \leqq \mu < \infty$, then (21) would be used with $\mu=\mu_0$. The left member of (21) and of (22) is a function of the observations $X_1, X_2 \ldots, X_n$ known as Student's t[20] with $n-1$ degrees of freedom. Tables giving certain of the percentiles of the distribution of this function for various values of n are widely available. Let $t_{n,p}$ denote the $100p$ percent point of the distribution of Student's t with n degrees of freedom; that is, the point at which the ordinate of the frequency curve of this distribution bounds a right-hand tail area exactly equal to p. Then if the producer's risk is to be fixed, the formula for determining c is simply

$$c = \pm t_{n-1,\alpha}$$

with the sign chosen so as to place the region of rejection properly in accordance with the obvious requirements of the hypothesis. If the consumer's risk is to be fixed, the formula becomes

$$c = \pm t_{n-1,\beta}$$

Numerical examples and further details will be found in [9].

The procedure given above is applicable for any proposed sample size n and does not fix the value of n. Perhaps the most satisfactory method of selecting n is by inspection of a family of power curves for the Student's t test drawn for various different values of σ and n.[21] If the use of alternative hypotheses seem to be appropriate, then these curves will be found to give in convenient form, for each n, the various values of the consumer's risk (assuming that the producer's risk was fixed in setting up the plan) which correspond to the particular values of σ, which are likely to be encountered in the applications.

It is perhaps a little discouraging to notice in this connection that if a sampling plan is set up by the formulas given in case 1, or case 2 in the appendix, and if the Student's t plan with, say, fixed producer's risk is set up for the same defi-

[20] See, for example, [9], p. 14. A table of the distribution of t is given on p. 167 of [9].
[21] See [8].

nition of good quality, the same producer's risk, and the same sample size, then the consumer's risk for a given definition of bad quality and a given upper bound for σ is considerably greater in the case of the Student's t plan than in the case of the other plan. (Of course, this consumer's risk in the Student's t plan can be reduced by increasing the sample size beyond that of the other plan). This fact suggests that the use of the Student's t approach in acceptance sampling work should be limited to cases in which it is essential at all costs to maintain either a constant producer's risk or a constant consumer's risk. But in practice, the exact validity of the underlying assumption of normality is always in doubt, and the method of sampling is seldom truly random, so the concept of a constant risk is in reality quite a nebulous one. *There is thus some basis for a recommendation that the Student's* t *approach should never be used at all in acceptance sampling work.*

The dependence on σ of the power curve of tests relating to μ in the normal case can be eliminated by the device of using a two-sample test, or double-sampling plan, in which the first sample serves the purpose of exploring the situation as regards σ, and the sample size of the second sample is adjusted accordingly. If the assumption of normality could be removed, such double-sampling plans might be of considerable utility in acceptance sampling work in which a variety of unique aggregates of uncertain pedigree and composition are to be sentenced by sampling in the complete absense of any history on which to base estimated upper bounds for σ. But as far as the writer is aware, there has been no attempt to eliminate the normality assumption in the literature of such tests to date. The assumption of normality, or even approximate normality, in the situation just now described would seem to be rather dangerous.

If the normality assumption is admitted, then there are several ways of going about the problem of setting up the test. A particularly interesting and straightforward method is given by Stein in [14]. Although it is beyond the scope of this paper to discuss multiple or sequential sampling plans at any length, it seems worth while to give here the instructions for setting up an acceptance sampling plan for alternative hypotheses on the basis of the Stein theory. For the hypotheses of case 1 of the appendix, the steps are as follows in terms of the notation introduced in this section and previously:

(1) Select a random sample of size m from the aggregate, and calculate s^2 for this sample. (The sample size m is not prescribed by the conditions of the problem, and is left to the judgment of the person doing the sampling. A small m might result in a large total sample size if σ is large.)

(2) Find the numerical value of

$$n'=1+\frac{s^2(t_{m-1,\alpha}+t_{m-1,\beta})^2}{(\mu_0-\mu_1)^2}.$$

Let n be the total number of observations in the first and second samples. The value of n is to be taken as $[n']$ [22] or m, whichever is the larger. (This means that if $[n']>m$, it will be necessary to take a second sample of $[n']-m$ items, but if $[n]\leqq m$, no second sample is necessary.)

(3) The region of rejection is now defined to be

$$\frac{\bar{x}-\mu_0}{s/\sqrt{n}}<c, \tag{23}$$

where

$$c=-t_{m-1,\alpha} \text{ and } \bar{x}=\frac{\sum_1^n x_j}{n}.$$

(Note that s is computed only from the first sample of m items, but $\bar{x}$ is computed from the first and second samples combined, if there is a second sample.)

For the hypotheses of case 2 of the appendix, the sense of the inequality (23) and the sign of c is reversed; the instructions are otherwise unchanged.

The critical region defined by (23) is obviously a very complicated one, as n is a function of the first m observations. It is not known to the writer whether or in what sense the test so defined has optimum properties.

It is important to observe that the producer's and consumer's risks of a double-sampling plan such as the above one apply to *the entire double-sampling process* and not just to the second sample. In sampling a sequence of aggregates, it would not be legitimate to select the first sample of m only in the case of the first aggregate and then to use the value of s and n so derived for all the succeeding aggregates. The value of s must be determined separately for each aggregate of the sequence.

[22] The symbol $[n']$ means largest integer in n'.

XI. Quality Measured Simultaneously by Several Parameters

It occurs frequently that the quality of an item must be measured by a number of variables that may be correlated with varying degrees of intensity.

If the quality of the sampled aggregate is to be measured by the mean value of two or more variables, and if these variables may be considered on engineering grounds to be statistically independent, then it is very easy to generalize the theory of the single-parameter case. For example, suppose that aggregate quality is to be measured by the three arithmetic mean values μ_1, μ_2, μ_3, and that the hypothesis H_0 of good quality has the form

$$a_1 \leqq \mu_1 < \infty, \; a_2 \leqq \mu_2 < \infty, \; a_3 \leqq \mu_3 < \infty.$$

Let α_1, α_2, and α_3 be the producer's risks associated individually and respectively with these three inequalities; that is, the producer's risks for sampling plans designed to control the means, one at a time. If $\bar{x}_1$, $\bar{x}_2$, and $\bar{x}_3$ are the three respective sample means, then an efficient test of H_0 simply consists in requiring rejection if one or more of the $\bar{x}$'s fall below acceptance numbers determined individually for the three quality characteristics by the methods used for the single-variate case. From this it can be inferred that the total producer's risk of the entire sampling plan (that is, the maximum probability of rejecting H_0 if H_0 is true) is given by

$$1-(1-\alpha_1)(1-\alpha_2)(1-\alpha_3)=\alpha_1+\alpha_2+\alpha_3-\alpha_1\alpha_2-\alpha_1\alpha_3-\alpha_2\alpha_3+\alpha_1\alpha_2\alpha_3.$$

If each of these risks is small, say less than 0.05, as is usually the case in practice, the product terms can safely be ignored, and the total producer's risk becomes simply the sum of the three individual risks.

On the other hand, the problem of the consumer's risk requires a different approach. The aggregate would usually be considered to be of bad quality if *any one* of the three parameters μ_1, μ_2, μ_3 lies in an interval that is associated with bad quality for that parameter; that is, an aggregate is generally considered unacceptable if it is bad in any one respect. Under these conditions the total consumer's risk of the sampling plan would be simply equal to the largest of the individual consumer's risks associated with sampling plans designed to control the means one at a time. Of course this point of view is somewhat arbitrary, and perhaps there would be circumstances in which an aggregate would not be rejected unless an inference could be made that it was bad on two or more counts. In this case the total consumer's risk would be the largest of the products of pairs of individual consumer's risks.

The special case of three parameters so far considered can immediately be generalized to a higher number of independent quality characteristics. The general result is that the producer's risks for individual quality characteristics are always added to get the total producer's risk, but the consumer's risks for each of the quality characteristics must ordinarily be considered separately.

A similar situation occurs when aggregate quality is simultaneously measured by the arithmetic mean of a quality characteristic and by the standard deviation. If a normally distributed aggregate can be assumed, the customary sample estimates of these parameters are statistically independent. (This is true asymptotically for any aggregate.) Thus if sampling plans are set up individually for these quality measures in accordance with the formulas in the appendix, the combined producer's risk will be approximately the sum of the two producer's risks, but the consumer's risks cannot be combined.

When the quality characteristics are correlated, the problem becomes far more complicated, both from the engineering and from the computational points of view. The definition of what is good and what is bad quality is often harder to decide upon in this case. For instance, a lot of plywood that is low on shear strength but high on percentage wood failure in sheared specimens might be just as acceptable as a lot that is very high on shear strength but very low on percentage wood failure in sheared specimens. Thus in the case of correlation, the definition of good quality may correspond in some cases to a spherical or ellipsoidal region of the space whose coordinates are the population parameters representing quality characteristics.[23] The formulation of the "best" test of the hypothesis of good quality in such cases and the computation of the risks may involve a

[23] It is conceivable that such a definition of good quality might occasionally be considered appropriate even in the case of independent characteristics.

755528—47——7

number of difficulties. Even for the hypothesis H_0 considered above, the computation of the producer's risk in a given case would involve a rather laborious mechanical quadrature. However, the formula given above for the case of independent variables provides an upper bound that often suffices in practical applications.

XII. Lot Acceptance and Process Acceptance

It has been emphasized several times in the foregoing discussion that the theory of inferential statistical sampling plans always involves the assumption that the sample is selected in such a way that the distribution of the sample observations bears a known, calculable relation to the unknown frequency distribution of the aggregate. In particular, it will be recalled that the validity of the formulas developed in sections VIII to X is dependent upon a method of sampling that will insure that the observations are statistically independent and that each has a probability distribution identical with the frequency distribution of the aggregate.

When such a requirement is considered from the operational point of view, it becomes necessary to distinguish between the case in which the sampled aggregate is all physically present and complete in itself at the time of sampling, and that in which the aggregate is viewed as a whole stream of items, some of which have yet to be manufactured. In other words, is it the purpose of the sampling plan to sentence a finite "lot" or delivery already produced, or the entire process? It has been customary to call the plan a "lot acceptance sampling plan "in the first case. The second case might be called "process acceptance sampling," and is one of the fundamental techniques of statistical quality control.

In lot acceptance sampling work, it is always possible, though often not practicable, by the use of a table of random numbers or some such device, to obtain more or less complete theoretical conformance with the requirement of "randomness," which is basic to the formulas of sections VIII to X. In process acceptance work, the difficulties in securing randomness with respect to the entire process are much more formidable. In addition, the viewpoint is broader, and economic and psychological considerations may dictate that it is more important to avoid erroneous rejection of a good process (and consequent shutdown of machinery) than it is to avoid temporary erroneous acceptance of a substandard process.

These considerations suggest that a rather frequently repeated statistical test of low nominal power and with a very low nominal producer's risk might be more suitable for process control than a less frequently repeated test of considerably higher power. Such an approach was advocated by Shewhart [15] and has now become the accepted sampling technic for statistical process control. An essential feature of the Shewhart system consists in plotting the sample data on a "control chart." This graphical device has been found useful in bringing about a condition in which the sequence of observations on successive items produced by the process has the characteristics of a sequence of observations on items successively drawn "at random" (in the sense of section VIII) from a finite aggregate. Shewhart calls this condition that of "statistical control" and has pointed out a number of benefits that it entails.

When a process is known to be in statistical control, a fairly extensive history of the production process is necessarily at hand. If this history indicates that the distribution of the aggregate has remained for some time within a subclass of distributions identifiable with good quality, then it is the viewpoint of many experts in the field of statistical quality control that the weight of such historical data should be ordinarily sufficient to permit the continual acceptance of lots without the use of a more powerful (and more expensive) lot acceptance sampling plan.

Such a viewpoint clearly rests on experience rather than on mathematical theory. If the purchaser does not have his own inspectors in the plant of the manufacturer, he may very well feel that he needs some sort of definite lot-by-lot protection against lack of real control or shifts in the distribution of a statistically controlled process. Moreover, each lot produced by a process in statistical control is itself a random sample from the aggregate consisting of the process, so a poor lot will occasionally be produced by a good process through sampling fluctuations.

Thus it is by no means irrational under some circumstances to employ a lot acceptance sampling plan on lots produced by a process supposedly in a state of statistical control. If the decision is

made to do this, the *historical data* concerning the control of the process, when available, will often provide safe estimates of nuisance parameters such as σ in the case of hypotheses concerning the mean and also will provide appraisals of the applicability of assumptions concerning the general form of the distribution of the lot.

It is interesting to notice that if a lot acceptance sampling plan is set up in connection with a controlled process, with the ultimate aim of accepting or rejecting the process, then the chances of making an erroneous decision are likely to be about the same for the individual lot and for the process. More precisely, if a sampling plan involving a hypothesis concerning, say, the mean μ of the sampled aggregate (assumed to be normally distributed), is set up for a pair of hypotheses similar to those of case 1 of the appendix, with the process playing the role of the sampled aggregate, then the upper limit of the conditional probability of rejection for a lot with mean $\overline{X}$ satisfying the condition $\overline{X} \geqq \mu_0$ is less than (but only slightly less than) α, and the upper limit of the conditional probability of acceptance for a lot with $\overline{X} \leqq \mu_1$ is less than (but only slightly less than) β. This situation arises from the fact that for any given μ, the "unconditional" marginal distribution of $\overline{x}$ (the sample mean) is normal with mean μ and standard deviation $\sigma/\sqrt{n}$, while for a given fixed value of $\overline{X}$, the conditional distribution of $\overline{x}$ (which is incidentally independent of μ) is normal with mean $\overline{X}$ and standard deviation $(\sigma/\sqrt{n})$ $(\sqrt{(N-1)/N})$ where N is the lot size), which is only very slightly less than $\sigma/\sqrt{n}$ for reasonably large values of N. Similar considerations apply to sampling plans involving the standard deviation.

Thus under the hypothesis of control, if the mean of the process is in a region associated with good quality, the occasional "bad" lot that is produced by an accident of chance itself has a high probability of being rejected by the sampling plan. The probability of accidentally accepting a bad lot under such circumstances is a little less than the probability, multiplied by β, that a bad lot will be produced. This probability of accidentally accepting a bad lot, for $N > n$, is infinitesimal for the pairs of hypothesis considered in this paper; for example, in the case of hypotheses concerning the mean, the region for $\overline{X}$ identified with bad lies well beyond the acceptance number c, which itself lies in the extreme tail of the distribution of a sample of n.

This suggests that when there is a certain amount of assurance that a state of statistical control exists, a "cheap" lot acceptance sampling plan with low power might be the reasonable one to use. For given definitions of good and bad quality, when statistical control is *not* assumed, it is customary to use consumer's and producer's risks of the order of 1 to 10 percent. It is suggested that when statistical control can be assumed, a satisfactory plan will result if a consumer's risk of 50 percent is used in conjunction with the value of the producer's risk, and the definitions of good and bad quality, that would have been chosen if no control were assumed.

XIII. Random Remarks on the Applications

In practice, the control of dispersion in lot acceptance sampling work is still usually handled by the method of imposing upper and lower tolerances on the variable in question, which are then enforced by sampling by attributes. However, the writer has seen a number of successful applications of variable technics involving the sample range or standard deviation. It seems probable that as statistical quality control methodology becomes more familiar to inspectors and factory personnel, the variables approach to the dispersion problem will be given proper emphasis in cases where it is not the position of the distribution, but its "width" that must be controlled. A case in point is to be found in recent issues of Navy Department specifications for certain types of high-pressure cylinders. The range of tensile strength in the steel in individual cylinders, which affects the weldability properties of the metal, is controlled in these specifications by means of a variables sampling plan.

In the case of quality measured by the mean, the situation as regards the present popularity of the variables approach is quite different. Acceptance sampling plans involving the use of the sample mean are very frequently met with in practice. Needless to say, they are usually set up on the assumption that the sample is an exact replica of the lot, and are then defended by their authors on the basis that "years of experience" have proved them to be satisfactory.

755528—47——8

It is easy to see why this type of approach actually does not often lead to trouble. The fact is that variables sampling plans involving the mean are usually applied in cases where there is considerable flexibility in the establishing of the definition of good and bad quality. By far the most frequent applications of this sampling technic are to cases in which the main problem is to see that a purchaser receives full measure under a contract or order. The replica-of-the-lot theory of sampling places the acceptance number for the sample mean squarely on the value of the mean agreed upon for the total amount of material under contract, and the contractor usually has sense enough to avoid rejections caused by sampling fluctuations by keeping his lot mean a little above the acceptance number. The purchaser, in his turn, has unconsciously established for himself a definition of a bad lot; but he seldom really knows exactly what is really in a given delivery, and a single underweight lot accidentally accepted through sampling errors will probably never be detected.

Closely related to the problem of full measure is the problem of the average analysis. In the purchase of certain commercial chemicals where complete homogeneity is not necessary, an average analysis is all that is needed to insure that the purchaser is getting his money's worth. This situation also obtains in the purchase of many metal items, such as pots and pans. In the same category are service life lot acceptance problems, in which the item is easily replaced from stockroom supplies if it wears out in use. Lamp bulbs, welding rods, searchlight carbons, and abrasives are all examples of materials which, under mass procurement conditions, give rise to inspection problems in which the estimation of the average service life of the items in each lot (perhaps by means of an accelerated life test) is the really important problem. In all these cases, the definition of what is good and what is bad quality is never really clear-cut. It is safe to say in general that empirically set up sampling plans using variables are not as likely to be really disadvantageous or pernicious from the purchaser's point of view as are certain empirical attributes sampling plans, such as the familiar "take one in each lot of 250" plan.

To be sure, the writer has seen some exceptions to this statement, which have occurred when specification writers in attempting to extrapolate from "years of experience" unconsciously have run afoul of the laws of probability. An example that comes to mind is a case in which the *mean* of a fairly large preliminary sample of observations on the breaking strength of glass portlights was taken by a specification writer as the *acceptance number* for the mean of a sample of four portlights selected from each lot. In another case, the mean of the buoyancy values of a few experimental kapok-stuffed life preserver pads was reduced arbitrarily by 1 pound and then used in the ensuing specification as a lower lot tolerance limit to be strictly enforced by attributes sampling. Because the standard deviation of most contractors for pad-stuffing was about ¾ pound (as could have been roughly inferred from the original data), trouble struck quickly when the specification hit the street.

These examples are admittedly somewhat extreme. It is not on the basis of such incidents that the scientific approach to variable acceptance sampling plans should be recommended, but rather on the basis of the fact that such an approach pays off in the long run. As in so many other fields of application of the modern theory of statistical inference, one can always do better on the average by using the theory than by not using it. And once in a while, a spectacular saving results or a serious error is averted, which alone justifies the extra trouble involved in basing the work on sound theory.

The chief obstacles (other than the difficulty of training personnel involved) to overcome in applying statistical theory to acceptance sampling by variables are that the sample selection should be truly random and that prior data on the form of the distribution, or at least on its higher moments, must generally be obtained. The problem of randomness has already been mentioned briefly in section XII and will not be further elaborated on here. The question of estimating bounds for higher moments is worth discussing here in a little more detail, at least in the important case in which the hypothesis concerns the mean and limits must be assigned to the dispersion of the sampled aggregate.

In some cases, the dispersion of the population is not due so much to nonhomogeneity of material as to lack of precision or reproducibility of a test, and once the testing error has been determined from a preliminary sample, it is likely to remain

rather constant. Examples of cases in which the lot, if properly defined, is likely to be rather uniform but the test (or tests) may exhibit considerable variability, are soap, impregnite, and various commercial chemicals produced by wet mix processes. In the case of chemical compounds blended by dry mix processes and in the case of metals produced by continuous melting furnaces, the composition of a lot is likely to change slowly and uniformly from the first portion poured to the last portion, and if it were not for the error of the test, purposively selected samples from the first and last portion should "bracket" the lot. In these cases, it is not the number of sample items, but rather the number of repetitions of a test or an analysis that must be calculated on the basis of the known precision of the analysis.

But in the majority of cases, it is the material rather than the test that is variable, and predictions as to future dispersions must be made on the basis of familiarity with the engineering process, intuition, and a certain amount of historical evidence.

In the first place, purely engineering considerations that limit the range of the variable under observation will sometimes yield useful upper bounds for the standard deviation of the sampled aggregate. In the gaging of the inside diameter of a bearing, for example, it may be known that the range of measurements in a lot will almost surely be ± 0.001 inch, from which a reasonable bound for the standard deviation can be inferred. In the production of fiber glass, the range of diameters of the fibers may be pretty well established by the process; in the winding of metallic-asbestos spiral-wound gaskets, the length of the strips of asbestos used will insure that the number of plies will fall within definite limits. An interesting case is furnished by resistance wire, which is subjected to average requirements on diameter mainly so that a piece of wire with a given resistance will not take up too much or too little room when wound on a mandrel. Unless different sizes of wire accidentally get mixed in a lot (which does sometimes happen but which would probably be detected by the resistance test), it can safely be inferred that the range of diameters in a lot will be restricted by the dies so that it lies between certain known limits.

XIV References

[1] H. F. Dodge and H. G. Romig, The Bell System Tech. J. **8,** 613 (1929).

[2] L. E. Simon, An engineers' manual of statistical methods (John Wiley & Sons, Inc., New York, N. Y., 1941).

[3] S. S. Wilks, Mathematical statistics (Princeton University Press, Princeton, N. J., 1944).

[4] A. Wald, J. Am. Stat. Assn. **40,** 277 (1940).

[5] J. Neyman and E. S. Pearson, Phil. Trans. Roy. Soc. (London) [A], **231,** 289 (1933).

[6] J. H. Curtiss, Ann. Math. Stat. **17,** 62 (1946).

[7] The Navy sampling inspection manual, chapter D4.00.00 in the administrative manual for the Materials Inspection Service, U. S. N., 1946.

[8] Charles D. Ferris, Frank E. Grubbs, and Chalmers L. Weaver, Ann. Math. Stat. **17,** 178 (1946).

[9] Harry Freeman, Industrial statistics (John Wiley & Sons, Inc., New York, N. Y., 1942).

[10] H. Cramér, Mathematical methods of statistics (Princeton University Press, Princeton, N. J., 1946).

[11] B. H. Camp, Bul. Am. Math. Soc. **28,** 427 (1922).

[12] J. V. Uspensky, Introduction to mathematical probability (McGraw-Hill Book Co., Inc., New York, N. Y., 1937).

[13] C. C. Craig, Ann. Math. Stat. **4,** 94 (1933).

[14] C. Stein, Ann. of Math. Stat. **16,** 243 (1945).

[15] W. A. Shewhart, Economic control of quality of manufactured product (D. Van Nostrand Co., Inc., New York, N. Y., 1931).

XV. Appendix

Single-Sample Formulas for Acceptance Sampling by Variables

1. Notation

X = a variable which measures a quality characteristic of an individual item

μ = the arithmetic mean of the frequency distribution of X in an aggregate of the items

σ = the standard deviation of the distribution of X in an aggregate of the items

n = sample size (number of items in a sample)

$\bar{x} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$, where

$X_1, X_2, \ldots, X_n$ are the observations on the sample items

$$s^2 = \frac{(X_1 - \bar{x})^2 + (X_2 - \bar{x})^2 + \ldots + (X_n - \bar{x})^2}{n-1}$$

c = acceptance number

α = producer's risk (maximum probability of rejecting an aggregate of good or acceptable quality)

β = consumer's risk (maximum probability of accepting an aggregate of bad or unacceptable quality)

k_p = defined by the equation

$$p = \int_{k_p}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

2. Pairs of Alternative Hypotheses

Case number	H_0		H_1	
	Definition of good quality	A priori information concerning parameters	Definition of bad quality	A priori information concerning parameters
1	$\mu \geqq \mu_0$	$\sigma \leqq \sigma_0$	$\mu \leqq \mu_1 < \mu_0$	$\sigma \leqq \sigma_1$
2	$\mu \leqq \mu_0$	$\sigma \leqq \sigma_0$	$\mu \geqq \mu_1 > \mu_0$	$\sigma \leqq \sigma_1$
3	$\mu \geqq \mu_0$	$\frac{\sigma}{\mu} \leqq \alpha_0$	$\mu \leqq \mu_1 < \mu_0$	$\frac{\sigma}{\mu} \leqq \alpha_1$
4	$\mu \leqq \mu_0$	$\frac{\sigma}{\mu} \leqq \alpha_0$	$\mu \geqq \mu_1 > \mu_0$	$\frac{\sigma}{\mu} \leqq \alpha_1$
5	$\mu \geqq \mu_0$	$\sigma \leqq \sigma_0$	$\mu \leqq \mu_1 < \mu_0$	$\sigma \leqq \sigma_1$
6	$\mu \leqq \mu_0$	$\sigma \leqq \sigma_0$	$\mu \geqq \mu_1 > \mu_0$	$\sigma \leqq \sigma_1$
7	$\sigma \leqq \sigma_0$		$\sigma \geqq \sigma_1 > \sigma_0$	
8	$\mu_0 \leqq \mu \leqq \mu_0'$	$\sigma \leqq \sigma_0$	$\mu \leqq \mu_1 < \mu_0$ or $\mu \geqq \mu_1' > \mu_0'$	$\sigma \leqq \sigma_1$

3. Single-Sample Formulas

(Random sampling)

Case number	Assumed form of distribution of aggregate	n	c	Region of rejection	Reference to notes
1	Normal	$\left(\frac{k_\beta \sigma_1 + k_\alpha \sigma_0}{\mu_1 - \mu_0}\right)^2$	$\frac{k_\beta \mu_0 \sigma_1 + k_\alpha \mu_1 \sigma_0}{k_\beta \sigma_1 + k_\alpha \sigma_0}$	$\bar{x} < c$	(a)
2	do	$\left(\frac{k_\beta \sigma_1 + k_\alpha \sigma_0}{\mu_1 - \mu_0}\right)^2$	$\frac{k_\beta \mu_0 \sigma_1 + k_\alpha \mu_1 \sigma_0}{k_\beta \sigma_1 + k_\alpha \sigma_0}$	$\bar{x} > c$	(a)
3	do	$\left(\frac{k_\beta a_1 \mu_1 + k_\alpha a_0 \mu_0}{\mu_1 - \mu_0}\right)^2$	$\mu_0 \mu_1 \frac{k_\beta a_1 + k_\alpha a_0}{k_\beta a_1 \mu_1 + k_\alpha a_0 \mu_0}$	$\bar{x} < c$	(a)
4	do	$\left(\frac{k_\beta a_1 \mu_1 + k_\alpha a_0 \mu_0}{\mu_1 - \mu_0}\right)^2$	$\mu_0 \mu_1 \frac{k_\beta a_1 + k_\alpha a_0}{k_\beta a_1 \mu_1 + k_\alpha a_0 \mu_0}$	$\bar{x} > c$	(a)
5	Lies entirely in finite interval $(\mu - \Delta, \mu + \Delta)$	$-18.42(\log_{10} \alpha) \frac{\sigma_0^2 + \frac{\Delta}{6}\lvert\mu_0 - \mu_1\rvert}{(\mu_0 - \mu_1)^2}$	$\frac{\mu_0 + \mu_1}{2}$	$\bar{x} < c$	(a), (b)
6	do	$-18.42(\log_{10} \alpha) \frac{\sigma_0^2 + \frac{\Delta}{6}\lvert\mu_0 - \mu_1\rvert}{(\mu_0 - \mu_1)^2}$	$\frac{\mu_0 + \mu_1}{2}$	$\bar{x} > c$	(a), (b)
7	Normal	$1 + \frac{1}{2}\left(\frac{k_\beta \sigma_1 + k_\alpha \sigma_0}{\sigma_1 - \sigma_0}\right)^2$	$\sigma_0 \sigma_1 \left(\frac{k_\beta + k_\alpha}{k_\beta \sigma_1 + k_\alpha \sigma_0}\right)$	$s > c$	(c)

4. Notes

(a) In the case of the "two-sided alternative" hypothesis, identified as case 8 in the list of pairs of alternative hypotheses, calculate sampling plan for each side separately by use of cases 1 and 2, or 3 and 4, or 5 and 6, as applicable, and use larger value of n for the final sampling plan. Final producer's risk will be less than or equal to the sum of the two producer's risks for each side.

(b) These formulas apply only when $\alpha = \beta$. In practice it is often easier to estimate Δ than σ_0, and it may safely be assumed that $\sigma < \Delta/3$. This leads to the following useful alternative formula for n:

$$-18.42\ (\log_{10} \alpha)\ \frac{\left(\frac{\Delta}{3}\right)^2 + \frac{\Delta}{6}\lvert\mu_0 - \mu_1\rvert}{(\mu_0 - \mu_1)^2}.$$

(c) Use these formulas only for $n > 30$. For $n \leqq 30$, set up sampling plan with the use of the χ^2 distribution.

Washington, November 22, 1946.

U. S. Department of Commerce
National Bureau of Standards

Research Paper RP1828
Volume 39, September 1947

Part of the Journal of Research of the National Bureau of Standards

Freezing Points of Cobalt and Nickel

By Milton S. VanDusen and Andrew I. Dahl

The freezing points of nickel and cobalt on the International Temperature Scale have been determined on samples from the same lots of metal used by Day and Sosman in their freezing-point determinations on the gas-thermometer scale. The freezing temperatures on the International Temperature scale are

Nickel ..... 1,455° ± 1° C.
Cobalt ..... 1,495° ± 1° C.

A value of the constant C_2 in Wien's (or Planck's) radiation law has been calculated on the basis of the observed ratios of brightness of black bodies at the freezing points of nickel and gold, cobalt and gold, and palladium and gold, and the values of the freezing temperatures of these metals on the thermodynamic scale as derived from the work of Day and Sosman. The average value of C_2 was computed to be 1.438 cm degrees.

I. Introduction

The International Temperature Scale [1] adopted in 1927 by the Seventh General Conference of Weights and Measures is based on a number of fixed points, the temperatures of which have been defined. The highest temperature so defined is that of the gold point, namely, 1,063° C. Above the gold point the temperature scale is extended on the basis of Wien's law of radiation. Any temperature, t, is determined by the ratio of intensity J_2 of monochromatic visible radiation of wavelength λ cm, emitted by a black body at temperature t, to the intensity J_1 of radiation of the same wavelength emitted by a black body at gold point, by means of the equation

$$\log_e \frac{J_2}{J_1} = \frac{C_2}{\lambda}\left(\frac{1}{1336} - \frac{1}{t+273}\right). \qquad (1)$$

The constant C_2 is taken as 1.432 cm degrees and must be such that $\lambda\ (t+273)$ is less than 0.3 cm degree.

The assignment of the value 1,063°C to the gold point is based largely upon the gas-thermometer work of Holborn and Day [2] and Day and Sosman.[3]

Day and Sosman extended the constant-volume nitrogen thermometer scale up to 1,550° C by obtaining the temperatures of various fixed points, including the freezing points of copper, nickel, cobalt, and palladium. The authors were very fortunate to have available for the present work samples of nickel and cobalt from the same lots of these materials that were used by Day and Sosman in their gas-thermometer work. Thus by obtaining the freezing points of these materials on the International Temperature Scale it was possible to compare the temperature scales at these points. The purity of the samples of nickel and cobalt was such as to give a positive significance to freezing-point determinations of these materials.

II. Previous Work on the Freezing Points of Nickel and Cobalt

With the exception of the determination of the nickel point by Wensel and Roeser,[4] practically no work has been done during the past 30 years on the freezing points of nickel and cobalt. Most of the early determinations were made by means of thermocouples, and the conversion of the reported values to the International Temperature Scale is uncertain. With the exception of the gas-thermometer work by Day and Sosman, the early work on nickel and cobalt is for the most part of historical interest only. The values for the freezing points of nickel and cobalt are summarized in table 1.

Where sufficient information was given, the reported values were adjusted to the International Temperature Scale.

[1] G. K. Burgess, BS J. Research **1**, 635 (1928) RP22.
[2] L. Holborn and A. Day, Am. J. Sci. **10**, 17 (1900).
[3] A. Day and R. B. Sosman, Carnegie Inst. Wash. Pub. 157 (1911).
[4] H. T. Wensel and Wm. F. Roeser, BS J. Research *5*; 1309 (1930) RP258.

TABLE 1.—*Summary of determinations of freezing points of nickel and cobalt*

Observer	Date	Method	Reported value	Adjusted value
		NICKEL		
			°C	°C
Holborn and Wien [a]	1895	Thermocouple	1,484	
Harker [b]	1905	do	1,427	
Copaux [c]	1905	do	1,470	
Burgess [d]	1907	Optical pyrometer	1,435	1,445
Day and Sosman [e]	1911	Gas thermometer	1,452	
Wensel and Roeser [f]	1930	Optical pyrometer	1,455	1,455
VanDusen and Dahl	1947	do	1,455	1,455
		COBALT		
Copaux [c]	1905	Thermocouple	1,530	1,526
Burgess [d]	1907	Optical pyrometer	1,464	1,475
Day and Sosman [e]	1911	Gas thermometer	1,490	
Raydt and Tammann [g]	1913	Thermocouple	1,480	1,484
Ruer and Kaneko [h]	1913	do	1,491	1,495
Burgess and Waltenberg [i]	1914	Microoptical pyrometer	1,478	1,482
VanDusen and Dahl	1947		1,495	1,495

[a] L. Holborn and W. Wien, Wied. Ann. **56**, 360 (1895).
[b] J. A. Harker, Proc. Royal Soc. (London) [A] **76**, 235 (1905).
[c] H. Copaux, Compt. rend. 140, 657 (1905).
[d] G. K. Burgess, Bul. BS **3**, 345 (1907) S62.
[e] A. Day and R. B. Sosman, Carnegie Inst. Wash. Pub. 157 (1911).
[f] H. T. Wensel and Wm. F. Roeser, BS J. Research **5**, 1309 (1930) RP 258.
[g] U. Raydt and G. Tammann, Z. anorg. Chem. **83**, 446 (1913).
[h] R. Ruer and K. Kaneko, Ferrum **11**, 5 (1913-14).
[i] G. K. Burgess and R. G. Waltenberg, Bul. BS **10**, 79 (1914) S205.

III. Materials

A detailed discussion giving the sources and the analyses of the metals used by Day and Sosman in their gas-thermometer work is given in their book, page 85. The analyses of the nickel and cobalt ingots made at the conclusion of the present work are shown in table 2.[5]

The compositions were essentially the same as those originally reported by Allen of the Geophysical Laboratory, with the exception of the iron content of the cobalt. A test for iron in a sample of cobalt that had not been used in the freezing-point determinations gave only 0.004 percent of iron. This indicated that the cobalt had become slightly contaminated with iron during drilling operations for the installation of sight tubes. This iron content, however, will not affect the freezing point of cobalt to any significant extent.

TABLE 2.—*Results of analyses*

Nickel		Cobalt	
Impurity	Amount	Impurity	Amount
	Percent		*Percent*
Pt	0.002	Pb	0.0014
Cu	.044	Cu	.016
Co	.106	Fe	.031
Fe	.004	S	.013
		Ni	.018

[5] Analyses made by R. K. Bell, of the Chemistry Division of this Bureau.

IV. Present Work

1. General Procedure

The so-called crucible method was followed in the freezing-point determinations. In this method a hollow enclosure or black body is immersed in a bath of the metal, and observations on the radiation from the black body are made with an optical pyrometer during the time the metal is freezing. In the present work the melting was done by means of a high-frequency induction furnace. This type of heating provides for a high degree of temperature uniformity throughout the sample and minimizes the probability of contamination of the sample by the furnace elements.

2. Crucibles

Previous use of beryllia and thoria as crucible and sight tube materials for melting pure metals led to their use in the present work. Thoria proved to be more satisfactory because of its greater mechanical strength. The beryllia and thoria crucibles and accessory parts were prepared by the general method described by Swanger and Caldwell.[6] The crucible and the sight tube were always of the same material.

3. Atmosphere

As both nickel and cobalt oxidize readily at high temperatures, it was necessary to melt the metals in either an inert or a reducing atmosphere. Atmospheres of hydrogen [7] and nitrogen (see footnote 3), as well as vacuum (see footnote 4) have been used by previous investigators. Helium was chosen for the present work as all available evidence indicates that the solubility of helium in either nickel or cobalt is less than that of hydrogen or nitrogen. The helium was sufficiently freed of oxygen to prevent any visible oxidation

[6] Wm. H. Swanger and F. R. Caldwell, BS J. Research **6**, 131 (1930) RP 327.
[7] G. K. Burgess and R. G. Waltenberg, Bul. BS **10**, 79 (1914) S205.

freezing point samples. Commercial helium was passed through copper chips heated at about 750° C to remove oxygen and then through magnesium perchlorate and phosphorus pentoxide to remove water vapor.

For comparison purposes, the freezing temperatures of nickel and cobalt were also measured in vacuum.

4. Arrangement of Apparatus

Figure 1 shows a cross-sectional view of the crucible assembly. The samples of nickel and cobalt were formed into the proper shape by melting them in beryllia crucibles of the size and shape to be used in the freezing-point determinations. Each sample weighed about 80 g.

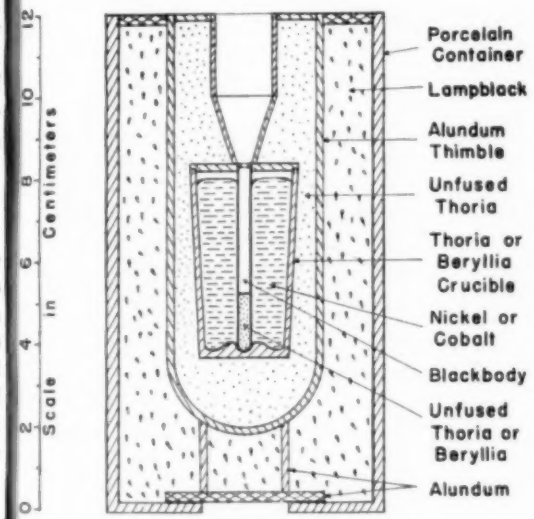


FIGURE 1. *Crucible assembly.*

An axial hole was drilled in each ingot to accommodate the sight tube. Each sample, together with its sight tube, was placed in a refractory-oxide crucible and mounted in a larger porcelain container, as shown in figure 1. The relatively coarse insulating material surrounding the crucible permitted the passage of gas around the sample, and the lampblack provided excellent heat insulation. The assembly shown in figure 1 was placed inside a Pyrex-glass enclosure, shown in figure 2, in which the desired atmosphere could be maintained. The tube was closed at the top with a Pyrex window sealed to the tube with propylene phthalate. The Pyrex tube fitted closely inside the water-cooled coil of the induction furnace.

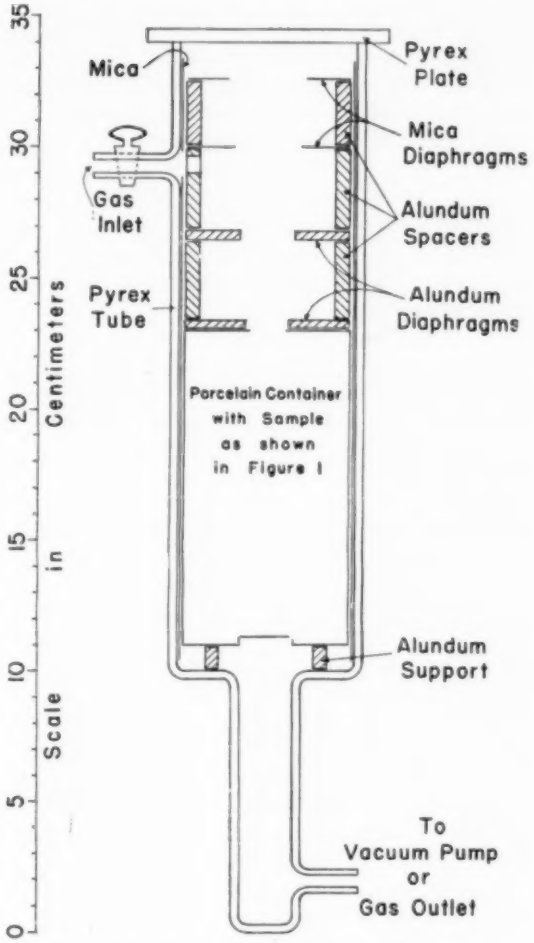


FIGURE 2. *Crucible assembly in Pyrex envelope.*

In the experiments employing a helium atmosphere, the gas was passed slowly through the Pyrex enclosure for about 16 hours before the heating was begun as well as during the actual heating. For the experiments in vacuum the pressure within the enclosure was maintained between 0.001 and 0.005 mm of Hg by means of a mercury diffusion pump. A liquid-air trap was located between the diffusion pump and the Pyrex tube to prevent mercury vapor from diffusing back into the tube.

5. Temperature Measurements

The optical pyrometer used was designed by Fairchild and Hoover.[5] It was fitted with a 45-degree total-reflecting for sighting into the black body. Two pyrometer lamps were used in the present work. For observations on the black body at the gold point, an assembly of the same dimensions as that shown in figure 1 was used. The difference was that only lamp-black insulation was used, and the crucible and sight tube were made of graphite.

The brightness ratio, J_2/J_1, was measured by utilizing a sector disk. The sector had two openings and was made by mounting steel strips on an aluminum disk. The edges of the strips were ground or lapped, straight and alined to be radial with the aid of a circular dividing engine. The transmission of the sector was 2.625 percent as computed from the measurement of the openings with a circular dividing engine. The sector transmission corresponded to a ratio of brightness from the freezing point of nickel to about 1,069° C for $\lambda_e=0.6533$ μ, the effective wavelength of the red screen over this temperature interval. Also, the sector transmission corresponded to a ratio of brightness from the freezing point of cobalt to about 1,093.5° C for $\lambda_e=0.6531$ μ, the effective wavelength of the red screen over the temperature interval in question. The values of 1,069° and 1,093.5° C were obtained from observations of the lamps at the gold point and a knowledge of the current temperature relation of the lamps over the temperature intervals involved.

In addition to using a sector disk as mentioned above, an absorption glass was used to obtain temperature measurements at the freezing points of the metals. This absorption glass has been calibrated repeatedly by comparison with sector disks over a period of about 20 years, and its calibration is considered well established.

A total of about 60 freezes on each metal were made, with both authors serving as observers. The average number of pyrometer settings during one freeze was seven. About 10 freezes on each metal were made with the samples in vacuum and the remainder with the samples in helium. No significant difference was noted. Both nickel and cobalt exhibited some undercooling. In a particular cobalt freeze, the metal was observed to undercool about 60 degrees before solidification started. It was possible to limit the magnitude of the undercooling by slightly jarring the furnace when the desired temperature was reached. An undercool of about 5 to 10 degrees was usually obtained.

As the transmission of the sector was such as to produce a match when the lamps were operated not exactly at the gold point (1,063° C) but at neighboring temperatures, the working equation becomes:

$$\log_e \frac{J_2}{J_0}=\log_e T=\frac{C_2}{\lambda_e}\left(\frac{1}{t_0+273}-\frac{1}{t_2+273}\right), \quad (2)$$

where T is the transmission of the sector.

If logarithms to the base 10 are used and C_2 expressed in micron degrees, the above equation may be written

$$\frac{1}{t_2+273}=\frac{1}{t_0+273}+\frac{\log_{10} T}{14320x0.43429}. \quad (3)$$

For a particular set of observations at the nickel point, $t_0=1068.9$° C. Substituting in eq 3, where $\lambda_e=0.6533$ and $T=0.02625$, we have $t_2=1455.2$° C. For a particular set of observations at the cobalt point, $t_0=1093.6$° C. Substituting in eq 3, where $\lambda_e=0.6531$ and $T=0.02625$, we have $t_2=1494.7$° C. A summary of the observational data is listed in table 3.

TABLE 3.—*Summary of determinations of the freezing points of nickel and cobalt*

Lamp number	Absorption device	Nickel		Cobalt	
		Number of freezes	Average value of freezing point	Number of freezes	Average value of freezing point
F-16	Glass No. 1	10	1,455.0	22	1,4
F-16	Sector No. 2	29	1,455.3	13	1,4
L-1	Glass No. 1	18	1,455.3	22	1,4
L-1	Sector No. 2	7	1,455.0	4	1,4
Weighted mean			1,455.2		1,4

An estimate of the errors entering into the measurement of the freezing temperatures of nickel and cobalt are given in table 4.

In view of the estimated uncertainties, the values of the freezing points of the samples of nickel and cobalt have been rounded to 1,455° C for nickel and 1,495° C for cobalt, with an uncertainty of 1° C.

[5] C. O. Fairchild and W. H. Hoover, J. Opt. Soc. Am. 7, 543 (1923).

TABLE 4.—*Estimate of uncertainties*

Source	Equivalent in deg C, with—	
	Sector	Glass
Transmission of sector	0.1	.....
Effective wavelength	.3	.....
Calibration of absorption glass	.....	0.8
Photometric matching	.3	.3
Departure from black-body conditions	.2	.2
Maximum, if all are of the same sign	0.9	1.3

The ratios of brightness, R, of black bodies at the freezing point of nickel and cobalt to a black body at the gold point may be obtained by substituting in the equation

$$\log_{10} R=\log_{10}\frac{J_t}{J_{Au}}=\frac{C_2\log_{10} e}{\lambda_e}\left(\frac{1}{1336}-\frac{1}{t-273}\right).$$

For nickel, where $\lambda_e=0.6533$, $R=41.40$. For cobalt, where $\lambda_e=0.6532$, $R=54.99$.

V. Calculation of the value of C_2 in Planck's law of radiation

Using the values of the freezing points of gold, nickel, and cobalt as determined by Day and Sosman, and the above values of the ratios of brightness, values of the constant C_2 in Planck's law of radiation can be calculated. For simplicity, Wien's law is used in the calculations as it yields the same values as those computed from Planck's law in the temperature range investigated. As the values of Day and Sosman for the temperatures of the various fixed points are on the constant-volume nitrogen scale, small corrections are necessary to convert them to the thermodynamic scale.[9] [10] Wien's law yields the following expression for C_2

$$C_2=\frac{\lambda_e\log_e\frac{J_t}{J_{Au}}}{\frac{1}{1336}-\frac{1}{t+273}}. \qquad (4)$$

In addition to the samples of nickel and cobalt used in the present work, Fairchild, Hoover, and Peters[11] have measured J_t/J_{Au}, for a sample of palladium that was compared with the palladium used by Day and Sosman and found to have the same freezing point to within 0.1° C. Table 5 lists the values of C_2 as derived from these points.

TABLE 5.—*Values of C_2 based on optical and gas thermometry*

Fixed point	Temperature [a] (Day and Sosman gas scale)	Thermodynamic scale	$\log_e \frac{J_t}{J_{Au}}\times 10^7$	C_2
	°C	°C		cm degrees
Au	1062.4	1062.6	.....	.....
Ni	1452.7	1453.1	2,433	1.4364
Co	1490.6	1491.0	2,618	1.439_8
Pd	1549.4	1549.8	2,879	1.438_2
			Mean	1.438_2

[a] These values are the unweighted means of the individual determinations of Day and Sosman and are slightly different from the weighted means reported by them.

Since the adoption of the International Temperature Scale in 1927 there has been considerable evidence to indicate that the value of 1.432 assigned to C_2 is low. The International Conference of Weights and Measures was scheduled to consider revision of the temperature scale in 1939 but failed to meet because of the unsettled political conditions. One of the proposed revisions was to increase the value of C_2 to 1.436 cm degrees. Since 1939 there has been evidence that the value should be still higher.

A value of the constant C_2 may be arrived at by a number of methods. Wensel [12] discusses these methods and summarizes the values of C_2 as obtained from the various sources.

The particular value of C_2 that was given the greatest weight was that derived from the atomic constants. In the most recent summary of the physical constants, Birge [13] gives a value of 1.4384_8 cm degrees for C_2.

WASHINGTON, June 6, 1947.

[9] E. Buckingham, Bul. BS 3, 237 (1907) S57.

[10] F. G. Keyes, Temperature, its measurement and control in science and industry, p. 45 (Reinhold Publishing Corp., New York, N. Y., 1941).

[11] C. O. Fairchild, W. H. Hoover, and M. F. Peters, BS J. Research 2, 931, (1929) RP65.

[12] H. T. Wensel Jour. Research NBS 22, p. 375 (1939) RP1189.

[13] R. T. Birge, Rev. Modern Phys. 13, 233, (1941).